TURING 图灵数学·统计学丛书 51

WILEY



An Introduction to Multivariate Statistical Analysis

多元统计分析导论

（第3版）



人民邮电出版社
POSTS & TELECOM PRESS

多元统计分析导论（第3版）
An Introduction to Multivariate Statistical Analysis

“……一本极佳的教材……全面阐述了多元统计分析中的数学理论……”

——*Clinical Chemistry*

“……依然是这个领域的权威著作，必定广受推崇……”

——*Short Book Reviews*

本书是一本优秀的多元统计分析教材，深受众多专家学者的好评。第1版于40多年前出版，其后陆续为众多学校采用为教材，经过不断修订和完善，成为多元统计方面公认的权威著作。本版系统阐述了多元统计中基本的论题，如假设检验、方差分析、回归分析、因子分析、主成分分析，并更大程度地运用极大似然方法，增加了相依性模式和图模型一章，系统介绍了椭球等高分布这一新的主题。

本书在我国统计领域已有广泛的影响，多年来是统计专业人士必读书籍，很多高校已将其指定为教材或主要参考书。

T. W. Anderson 1918年6月5日出生于美国明尼阿波利斯市，1945年获普林斯顿大学数学专业博士学位，后任教于芝加哥大学、哥伦比亚大学及斯坦福大学。美国科学院院士，数理统计学会、统计学会、经济协会、艺术与科学学会会士。Anderson教授一生获得过许多荣誉，且著述颇丰，在统计领域做出了卓越的贡献。



WILEY
www.wiley.com

图灵网站：www.turingbook.com 热线：(010)51095186
反馈/投稿/推荐信箱：contact@turingbook.com
有奖勘误：debug@turingbook.com

分类建议 数学/统计

人民邮电出版社网址：www.ptpress.com.cn



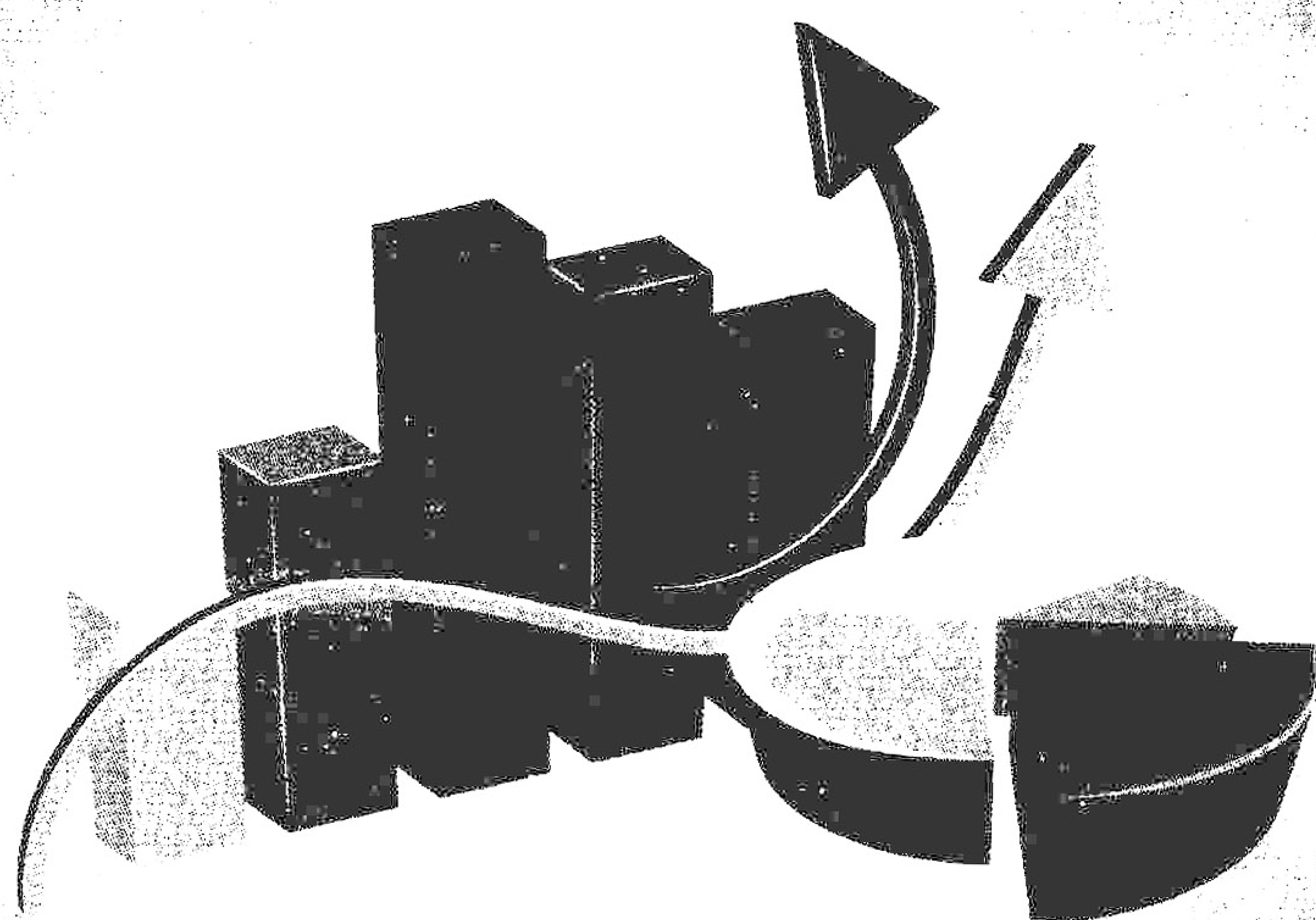


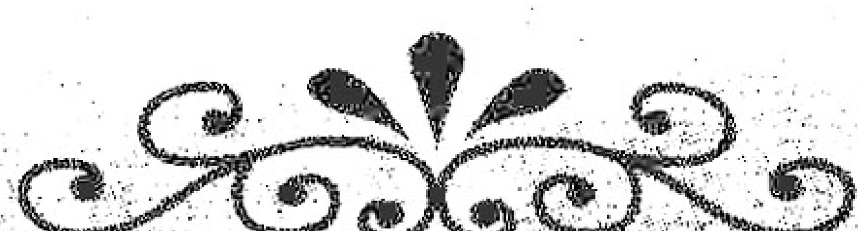


ISBN 978-7-115-24118-4
定价：89.00元

TURING 图灵数学·统计学丛书 51







An Introduction to Multivariate Statistical Analysis

多元统计分析导论

（第3版）

人民邮电出版社
北京

图书在版编目(CIP)数据

多元统计分析导论/(美)安德森(Anderson, T. W.)著; 张润楚等译. —北京: 人民邮电出版社, 2010.12
(图灵数学·统计学丛书)
书名原文: An Introduction to Multivariate Statistical Analysis, Third Edition
ISBN 978-7-115-24118-4

Ⅰ. ①多… Ⅱ. ①安… ②张… Ⅲ. ①多元分析: 统计分析 Ⅳ. ①0212.4

中国版本图书馆 CIP 数据核字(2010)第 202633 号

内 容 提 要

本书是世界知名统计学家的力作，主要内容有多元正态分布、方差分析、回归分析、因子分析、椭球等高分布、相依性模式、图模型. 附录中还列出了矩阵理论、Wilk 似然准则和其他常用检验的显著性水平的分位数.

本书在世界各高等学校中广为采用，是一本经典的多元统计分析课程的教材，也可供相关统计研究人员、应用多元统计的科技工作者参考.

图灵数学·统计学丛书

多元统计分析导论(第 3 版)

◆ 著 [美] T. W. Anderson
译 张润楚 程 轶 等
责任编辑 明永玲

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址: http://www.ptpress.com.cn
北京隆昌伟业印刷有限公司印刷

◆ 开本: 700×1000 1/16
印张: 35
字数: 703 千字 2010 年 12 月第 1 版
印数: 1–2 500 册 2010 年 12 月北京第 1 次印刷

著作权合同登记号 图字: 01-2007-5300 号

ISBN 978-7-115-24118-4

定价: 89.00 元

读者服务热线: (010)51095186 印装质量热线: (010) 67129223

反盗版热线: (010)67171154

版 权 声 明

译 者 序

作为人类认知世界最基础的共用科学之一的统计学, 像哲学、数学 (主要指纯粹数学) 一样, 应用非常广泛. 与哲学和数学所不同的是, 哲学通过对世界的总体看法和思考来认知世界, 数学通过抽象地研究数量、图形、符号等形式逻辑理论和方法来认知世界, 而统计学则通过对世界观察得到各种形态的信息资料 (称为数据) 进行分析, 研究有关的推断理论和方法来认知世界的. 哲学、数学和统计学的发展各自相对独立, 又相互依赖和支撑. 从历史上看, 统计学、数学和哲学思想的形成应该是相互交替进行的. 但归根结底, 人类的认知应该首先来源于对世界的观察. 由此看来，通过观察进行归纳和推断的统计思想应该更为基本. 哲学和数学是对观察结果的抽象, 它们经过发展所形成的各种形式逻辑和辩证逻辑使得推理更为严密, 因而反过来又会促进更为科学地观察和分析推断世界. 这点也许是为什么统计学思想的产生和发展虽然很早但现代统计学的形成却比较晚的原因之一.

我们有理由把统计学定义为一门通用科学, 它研究如何有效地观察世界, 得到信息 (各种形态的数据资料), 并由此来认知世界, 研究的问题包括如何收集、整理和分析数据以及进行推断. 统计学离不开对世界的观察, 也离不开数据和对数据的分析. 世界是复杂的, 存在各式各样的研究对象和各式各样的研究问题. 如果我们对任何一个研究对象的某个研究问题加以分析, 大都可以归结为对变量之间关系的研究. 因此通过得到的数据研究变量之间的关系, 就成为一个普遍关心的问题. 于是, 在统计学中, 多元统计分析就自然地成为最重要的研究分支之一.

对多元统计方法开始系统研究可以追溯到 20 世纪 20 年代甚至更早, 本书的作者 T. W. Anderson, 还有我国的许宝禄先生, 就是研究这类方法的先驱者. 本书的第 1 版出版于 1958 年, 它是对成书之前形成的各种多元统计理论和方法的最早的总结. 由于其理论的系统性和深刻性以及方法的多样性, 使其成为这个分支的第一个经典之作. 多元统计的理论和方法还在不断发展. 正如作者在本书的第 2 版和第 3 版的前言中所说, 在第 1 版之后, 多元统计理论和方法的许多方面都得到很快的发展. 现在, 国际上在发表这方面研究论文的同时, 出版了大量的多元统计分析的著作和教材, 许多都可以用来作为研究和应用的参考, 但我们认为, 这本由 T. W. Anderson 写的书仍然是我们想了解这个领域的首选之作, 尤其是在了解理论和方法方面, 它可以帮助我们了解对问题的开始提法和研究方向.

现在, 多元统计方法已经被广泛地用于各个领域, 包括金融学、经济学、社会学、生物学、医学、物理、化学、工学和农学等各学科, 还有环保、气象等各个领域.

希望这本书的中文版能促进我国在这个领域理论和应用的发展.

本书的翻译是由张润楚、程轶、庞芳、薛晓琳和曹琼合作完成的, 张润楚翻译了第 15 章和附录, 程轶、庞芳、薛晓琳和曹琼分别翻译了第 1~5 章、第 6~8 章、第 9~11 章、第 12~14 章，最后由程轶和张润楚进行审阅和协调统一. 郑贵鹏也参加了部分翻译工作并对书稿编辑和形成 CTX 文件做了不少工作, 图灵公司编辑们为本书的出版给予了大力支持, 包括仔细审校书稿, 我们特此表示感谢. 由于译者的学识水平以及时间所限, 本书的译文肯定会存在缺点甚至错误之处, 肯请读者给予批评指正.

张润楚

于南开大学和东北师范大学

2010 年 4 月

第 3 版前言

本书的前两版用来教授学生学习多元统计分析理论和方法已经快 40 年了. 同时该书在帮助广大统计学者增强对这个领域的了解和熟悉有关知识方面也发挥了作用. 自第二版出版以来, 多元分析在许多方向上已经得到很大发展和开拓. 在这个新版本里, 我不想论述所有扩大了的范围, 而只是选择对理论方法和综合理解特别有用的若干方面进行阐述.

前两个版本包含了只要用简单的计算器就可以完成的若干方法. 然而, 在 21 世纪的今天, 计算技术已经得到高度发展, 不可能用一本一般的数学理论书来包括所有的计算方法. 统计学某些方面也用到了再抽样技术这样的计算手段, 但这些并没有在这一版本里介绍.

多元统计处理相互间具有联系的变量. 这个版本我们仍用几章来讲述相关性的度量和独立性检验. 我们加进了新的一章, "相依性模式, 图模型". 所谓图模型, 是由一组顶点 (也叫结点) 连同一组连线 (也叫边) 构成的图, 这些顶点用于识别观测到的变量, 而这些连线用于反映变量间相依性关系. 这样, 图的代数理论便是路径分析和因果链研究的一个自然发展和产物. 一个图可以表达依时间先后或逻辑关系形成的序列, 也可以设想为一组变量是导致另一组变量变化的原因.

本版中, 我们系统介绍的一个新的论题是椭球等高分布. 我们知道, 由均值向量和协方差阵刻画的多元正态分布具有一个局限, 即变量的四阶矩由一阶矩和二阶矩就可以决定. 而椭球等高分布类放宽了这个限制, 在这个类中的分布密度也具有像正态密度一样的椭球型等高密度, 但是其变量的四阶矩多出一个自由度. 这个论题在书中新增的有关章节里给予详细阐述.

我们在第 12 章和第 13 章中详尽阐述的降秩回归提供了一个方法, 用以减少在一组变量对另外一组变量的回归中要估计的回归系数个数. 这一方法包含着在一个联立方程模型中一个方程的有限信息极大似然估计.

第 3 版的写作得益于前两版读者的建议和评论, 还有这一版本审稿人的意见. 除了前两个版本前言列出的读者以外, 我还要感谢 Michael Perlman 和 Kathy Richards, 他们为完成这个稿子给予了很大帮助.

T. W. Anderson

2003 年 2 月于加州斯坦福

第 2 版前言

自本书的第 1 版出版已经过去 26 年了. 在这段时间里, 多元统计分析得到很大发展, 特别在本书第 1 版所提到的那些方面. 这个新版本的目的是通过大量修订、改写和增添, 让这本书涵盖这个领域最新的发展. 这个版本仍保留原来的基本途径, 即对于每个主题由几种测量或刻画以及对它们性质的研究组成的统计方法在数学上的严格发展. 论题的一般梗概仍然保留.

在这个版本中, 从其他方面讨论了极大似然方法. 在均值向量和协方差阵的点估计中, 还介绍了关于某些损失函数比极大似然估计更好的估计, 如 Stein 估计和贝叶斯估计. 在假设检验方面, 似然比检验通过其他不变性方法处理得以补充. 给出了一些关于分布和渐近分布的新结果, 列表给出某些分位数. 对这些方法的性质, 比如功效函数、容许性、无偏性以及功效函数的单调性, 都进行了研究. 对均值和协方差的联立置信区间进行了研究. 用因子分析的一章代替了第 1 版中勾画零碎结果的那一章. 还介绍了一些新的论题, 包括联立方程模型和线性函数关系. 附加问题中介绍了进一步的结果.

本书不可能覆盖所有的有关材料, 但应该已经包括了那些我认为是最重要的内容. 关于直到 1966 年的论文和直到 1970 年的书的综合参考文献, 读者可以参看由 Anderson, Das Gupta 和 Styan (1972) 所著的*A Bibliography of Multivariate Statistical Analysis*. 进一步的参考文献可以在 Subrahmaniam 和 Subrahmaniam (1973) 著的*Multivariate Analysis: A Selected and Abstracted Bibliography, 1957-1972* 中查到.

我要感谢许多学生、同事和朋友, 他们提出了建议和帮助. 他们包括 Yasuo Amemiya, James Berger, Byoung-Seon Choi, Arthur Cohen, Margery Cruise, Somesh Das Gupta, Kai-Tai Fang, Gene Golub, Aaron Han, Takeshi Hayakawa, Jogi Henna, Huang Hsu, Fred Huffer, Mituaki Huzii, Jack Kiefer, Mark Knowles, Sue Leurgans, Alex McMillan, Masashi No, Ingram Olkin, Kartik Patel, Michael Perlman, Allen Sampson, Ashis Sen Gupta, Andrew Siegel, Charles Stein, Patrick Strout, Akimichi Takemura, Joe Verducci, Marlos Viana 和 Y. Yajima. 在本书写作过程中我还得到 Dorothy Anderson, Alice Lundin, Amy Schwartz 和 Pat Struse 的帮助. 特别感谢 Johanne Thiffault 和 George P. H. Styan 的关心. 本书的出版得到 Army Research Office, National Science Foundation, the Office of Naval Research 和 IBM Systems Research Institute 的资助.

附录 B 给出了 7 个分位数表, 以供进行检验方法使用. 表 1、表 5 和表 7 分别是 E. S. Pearson 和 H. O. Hartley 的著作*Biometrika Tables for Statisticians* 第二卷的表 47、表 50 和表 53; 表 2 是根据由 A. W. Davis 准备的和发表在*Biometrika* (1970a), *Annals of the Institute of Statistical Mathematics* (1970b) 和*Communications in Statistics, B. Simulation and Computation* (1980) 的 3 个表做出来的. 表 3 和表 4 分别是由 Ziro Yamauti (1977) 编辑和由 Japanese Standards Association 发表的*Concise Statistical Tables* 中的表 6.3 和表 6.4, 那是*Statistical Tables and Formulas with Computer Applications, JSA-1972* 的一个简明版本. 表 6 是 B. N. Nagarsenker 和 K. C. S. Pillai 以及 Aerospace Research Laboratories (1972) 发表的*The Distribution the Sphericity Test Criterion*, ARL 72-0154 的表 3. 感谢上述作者和出版者允许我在本书中使用这些表.

T. W. Anderson

1984 年 6 月于加州斯坦福

第 1 版前言

本书的主要目标是成为多元统计两学期课程的教材. 同时希望该书能作为那些不是学生的统计学者了解这个领域各方面的一个导引, 其他统计学家也可以把它用做参考书.

这本书曾以讲义笔记的形式在哥伦比亚大学的一门两学期研究生系列课中使用多年; 前 6 章构成第一学期的教学内容, 重点放在相关性理论方面. 假设读者熟悉一元统计的一般理论, 特别是了解基于一元正态分布的方法. 矩阵代数的知识也是必须具备的; 不过, 关于这方面的内容在附录中已经介绍了.

希望更基础和更重要的论题在这里得到处理, 虽然在某种程度上这种涵盖与个人喜好有关. 某些最近的前沿发展只是在后面的章节有简短的介绍.

本书在很大的程度上使用极大似然方法. 这引导出许多合理的结果, 且在某些情况下可以证明这些结果是最优的. 然而, 在许多情况下, 这些期望的或最优的结果还有缺陷.

在这本书稿不断完善的几年里, 大量的学生和同事给予了很大的帮助. 特别要提到的是 Allan Birnbaum, Harold Hotelling, Jacob Horowitz, Howard Levene, Ingram Olkin, Gobind Seth, Charles Stein 和 Henry Teicher. 还要感谢哥伦比亚大学的研究生数学统计学会在从讲义笔记到形成这本书的过程中所给予的协助. 本书的写作得到了 Office of Naval Research 的资助, 在这里一并致谢.

T. W. Anderson

1957 年 12 月于加州斯坦福

目 录

第1章 引 论

1.1 多元统计分析

多元统计分析的数据是由在若干个体或对象上的多组测量构成的. 这个样本数据, 可能是从某城市所有在校儿童中随机抽取的一些儿童的身高和体重; 也可能是一个测量集族, 比如两种鸢尾植物花瓣的长度和宽度以及萼片的长度和宽度; 或者又可能是若干学生在一系列心理测试中的得分.

可将一个个体上的多个测量写成一个列向量, 我们把这个向量整体看做是来自某多元总体或分布的一个观测. 当此个体是随机抽取时, 我们认为其测量对应的向量是一个随机向量, 它和总体有相同的分布或概率规则. 样本中所有个体的观测组成了向量样本, 将这些向量排成一行就形成了一个观测矩阵.① 从而那些需要分析的数据就表示为一个或若干个矩阵.

我们将会看到, 把每个观测向量当做欧氏空间里的一个点有助于直观地表示数据和理解方法, 其中观测向量的每一个坐标相应于一个测量或变量. 实际上, 统计分析首先要做的就是用图表示数据. 大多数统计学家仅限于二维作图, 即将观测向量的两个坐标依次表示在图中.

对于单变量分布, 其重要的特征是作为位置测度的均值和作为变异性测度的标准差. 同样, 对于单变量的样本, 其均值和标准差也是很重要的测度. 在多元分析中, 无论总体还是样本, 不同测量的均值和方差之间具有相应的相关性. 多元分析的一个重要内容就是研究不同变量间的相依性. 两个变量间的相依性可以用它们的协方差表示, 即它们分别与各自均值离差之积的平均. 用相应的标准差将协方差标准化, 就得到了相关系数, 它反映了这两个变量的相依程度. 因此对于多元变量, 可选择均值向量 (由单变量的均值组成) 和协方差阵 (由单变量的方差和两两变量的协方差组成) 作为一组描述统计量. 或者也可以选择均值向量、所有单变量的标准差和相关阵作为一组描述统计量集合. 这两组统计量包含相同的信息. 这些参量描述了一个总体或概率分布的位置、变异性和相依性. 多元正态分布就是完全由其均值向量和协方差阵决定的, 因此其样本均值向量和协方差阵就是一组充分统计量.

多元分析的基本内容是刻画和分析变量之间的相依性、变量集之间的相依性

① 当在纸上列表记录数据时, 会很自然地将一个个体的测量记为表上的一行, 因此一个个体相当于一个行向量. 但因为我们更喜欢用列向量进行代数运算, 所以用列向量来表示一个观测. (在实际中, 这些数据也可能记录在卡片、磁带或磁盘上.)

以及变量与变量集之间的相依性. 多重相关系数是相关系数的扩展, 它刻画了一个变量和一组变量之间的关系. 偏相关系数刻画了两个变量在除去其他相关变量影响后的相依程度. 从样本中得到的各种相关系数是用来估计总体分布中相应的相关系数. 本书将讨论独立假设检验, 以及从多元正态分布抽样的估计性质和检验方法.

多元总体的许多统计问题是单变量总体问题的直接类比, 解决这些问题的方法也是相似的. 例如, 在单变量情形下, 我们可能希望检验一个变量的均值为零的假设; 在多元情形下, 我们就可能希望检验若干个变量的均值组成的均值向量为零向量的假设. 前一个假设用到的学生 t 检验的类比是广义 T^2 检验. 单变量的方差分析也适用于观测值向量. 在回归分析中, 因变量可能是一个变量向量. 因此方差的比较就推广为协方差阵的比较.

单变量统计量的检验方法推广到多元情形下需要考虑变量间的相依性, 这些方法是不依赖坐标系的, 即这些方法相对于那些关于原假设不变的线性变换是不变的. 在一些问题中, 可能会有很多族检验是不变的, 这时则需要做出选择. 因此需要考虑检验的最优性质.

然而, 在另一些情况下, 选择一个坐标系是很重要的, 它可使变量具有所希望的统计性质. 有人会说它们具有正态分布或样本的固有特征. 这些和矩阵的典范型的代数问题有很大关联. 例如, 找一组变量的具有极大方差或极小方差的正规线性组合 (找主成分) 就相当于找一个坐标轴的旋转, 使得协方差阵变成对角形式. 另一个例子是刻画两个变量集之间的相依性 (找典型相关). 这些问题涉及不同矩阵的特征根和特征向量, 也要考虑相应样本量的统计性质.

一些统计问题产生于均值和协方差被限定的模型中. 因子分析可能是基于一个 (总体) 协方差阵, 这个协方差阵是一个正定对角矩阵与一个降秩的半正定矩阵之和; 线性结构关系也有相似的形式. 计量经济学中的联立方程组就是一个特殊模型的例子.

1.2 多元正态分布

本书讨论的统计方法是针对多元正态分布的, 但在抽样分布不是正态时, 许多方法也是有效的. 基于正态分布进行统计分析的一个主要原因是, 这个概率模型可以很好地近似很多抽样总体中连续测量的分布. 事实上, 大多数数据统计分析的方法和理论已经发展成熟. 一些数学家, 如 Adrian (1808), Laplace (1811), Plana (1813), Gauss (1823) 和 Bravais (1846) 研究了二元正态密度. 遗传学家 Francis Galton 在研究从一个父代上和一个子代上得到的成对测量时, 引入了相关、回归和同方差性的思想. [例如, 见 Galton (1889).] 他把多元正态分布理论表达为样本观测性质的一种推广.

Karl Pearson 和其他人继续发展了多元正态分布的理论, 并在遗传学、生物学和其他领域的研究中使用了不同类型的相关系数.[①] R. A. Fisher 进一步发展了应用于农学、植物学和人类学的方法, 包括分类问题的判别式函数. 在另一个方向上, 智力测验得分的分析引出了另一个理论, 它包括*因子分析*, 即基于正态分布的抽样理论. 统计学家发现, 在这些情况下以及在农业实验、工程问题、特殊的经济问题和其他领域里, 多元正态分布是足够近似于总体分布的, 从而, 基于正态模型的统计分析是合理的.

单变量正态分布的出现通常是由于所要研究的效应是若干独立随机效应之和. 类似地, 多元正态分布的出现通常是由于多重测量是一些较小的独立效应之和. 类似于单变量中心极限定理可以导出单变量正态分布, 多个变量的广义中心极限定理也可以导出多元正态分布.

基于正态分布的统计理论的优点是, 基于此的多元方法已被广泛地发展并可以进行有组织和系统性的研究. 这不仅是因为这些方法在实际应用中的需要, 还因为正态理论也适应于严格的数学处理. 这些恰当的分析方法主要运用了矩阵代数的标准运算; 可以得到许多相关的统计量的分布或至少可得到其特征; 并且在许多情况下, 可以推断方法的最优性质.

本书从多元正态分布的角度陈述所推断的问题, 在此基础上研究有效且通常最优的方法, 并且计算其显著性水平和置信水平. 这个途径具有阐述上的一致性和严密性, 但是, 就其本质而言, 显然它不能详尽地阐述全部多元统计分析的内容. 这些方法适用于许多非正态分布, 但是它们的充分性还值得怀疑. 大致来讲, 根据中心极限定理的运算, 关于均值的推断是稳健的, 但是关于协方差的推断对正态性是敏感的, 样本协方差的变化依赖分布的四阶矩.

正态方法关于阶数大于二的矩的不变性可以弱化到一大类椭球等高分布上. 在单变量情形中, 正态分布可由其均值和方差决定, 高阶矩以及峰度和长尾等性质都是均值和方差的函数. 同样, 在多元情形下, 均值向量和协方差阵, 或者说, 均值向量、方差向量及相关系数阵决定了分布的所有性质. 这个限制可以简化为一类很广泛的椭球等高分布. 这个类保持了相依性的结构, 且允许出现更一般的峰度和长尾. 对此的研究可以引出更多稳健方法.

计算机技术的发展给多元统计分析带来了多方面的改革. 在单变量统计中, 现代计算机支持用再抽样的方法去计算观测的变化和结果的显著性, 比如自助法和交叉核实法. 这些方法论弱化了对分位数表的依赖, 同时也消减了一些正态分布的限制.

非参技术适用于对潜在分布一无所知的情况. 由于篇幅所限, 本书不涉及这方

① 关于相关性思想发展的详细研究, 见 Walker (1931).

面的内容, 也不讨论数据分析的其他问题, 例如离群值的处理和变换变量为近似正态的和同方差的.

现代计算机设备的有效利用使得分析大型数据集成为现实. 这种计算能力促使多元方法应用于图像分析等新领域, 以及气象数据等数据的更有效的分析. 同时, 一些新的统计分析问题也随之而来, 比如参数或数据阵的稀疏性. 由于硬件和软件的迅猛发展以及编译程序需要特殊的知识, 在本书中将会对涉及计算的部分给出一些说明. 大多数的方法都有可利用的统计程序软件包.

第 2 章　多元正态分布

2.1　引　　言

本章我们讨论多元正态分布及其性质. 2.2 节介绍了多元分布的基本概念：多元密度函数、边缘分布 (也叫边际分布)、条件分布、期望值以及矩. 2.3 节给出了多元正态分布的定义, 其参数是表示此随机向量的均值向量、方差向量和协方差阵, 或者均值向量、方差向量和相关系数阵. 2.4 节证明了正态变量的线性组合还是正态分布的, 因此, 多元正态分布的边缘分布都是正态的. 在 2.5 节中, 我们可以看到多元正态分布的条件分布也是正态的, 其均值是那些条件变量的线性函数, 其中系数是回归系数. 方差、协方差和相关系数 (称为偏相关系数) 是常数. 多重相关系数是一个标量随机变量和其他随机变量线性组合之间的极大相关系数, 它衡量了一个变量和一个变量集之间的相依性. 正态分布的边缘分布和条件分布都是正态的, 这个事实使得在这一族分布上的处理连贯起来. 2.6 节讨论了特征函数、矩以及半不变量. 2.7 节给出了椭球等高分布, 并将正态分布的性质推广到这个更大的分布类上.

2.2　多元分布的概念

2.2.1　联合分布

本节我们介绍多变量联合分布的概念, 它将派生出变量子集的边缘分布和条件分布. 首先考虑两个 (实) 随机变量① X 和 Y. 依照这些随机变量定义的事件的概率可以通过关于*累积分布函数*(缩写为 cdf) 的运算得到. 对于任意一对实数 (x, y), cdf 的定义为

$$F(x,y)=\Pr\{X\leqslant x, Y\leqslant y\}. \tag{1}$$

我们对 $F(x,y)$ 绝对连续的情形感兴趣, 绝对连续意味着下面的偏导数几乎处处存在,

$$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y), \tag{2}$$

并且

$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\mathrm{d}u\mathrm{d}v. \tag{3}$$

① 在第 2 章中, 我们将分别用大写和小写字母来区分随机变量和现时运作 (running) 变量. 在后面的章节里, 由于其他复杂的符号, 我们可能不再遵守此约定.

非负函数 $f(x,y)$ 称为 X 和 Y 的密度. 随机变量对 (X,Y) 定义了平面上的一个随机点. (X,Y) 落在一个矩形里的概率是

$$\begin{aligned}&\Pr\{x \leqslant X \leqslant x+\Delta x, y \leqslant Y \leqslant y+\Delta y\} \qquad (4)\\ =&F(x+\Delta x, y+\Delta y)-F(x+\Delta x, y)-F(x, y+\Delta y)+F(x,y)\\ =&\int_y^{y+\Delta y}\int_x^{x+\Delta x} f(u,v)\mathrm{d}u\mathrm{d}v,\end{aligned}$$

其中 $\Delta x>0, \Delta y>0$. 随机点 (X,Y) 落在任意一个使得下面积分有定义的集合 E (即可测集 E) 中的概率是

$$\Pr\{(X,Y)\in E\}=\int\int_E f(x,y)\mathrm{d}x\mathrm{d}y. \qquad (5)$$

这来自积分的定义 [即 (4) 的有限和]. 如果 $f(x,y)$ 对每个变量都是连续的, 则概率元素 $f(x,y)\Delta y\Delta x$ 是 X 落在 x 和 $x+\Delta x$ 之间且 Y 落在 y 和 $y+\Delta y$ 之间的近似概率, 因为

$$\begin{aligned}\Pr\{x \leqslant X \leqslant x+\Delta x, y \leqslant Y \leqslant y+\Delta y\}&=\int_y^{y+\Delta y}\int_x^{x+\Delta x} f(u,v)\mathrm{d}u\mathrm{d}v \qquad (6)\\ &=f(x_0,y_0)\Delta x\Delta y,\end{aligned}$$

其中 x_0, y_0 $(x\leqslant x_0\leqslant x+\Delta x, y\leqslant y_0\leqslant y+\Delta y)$ 由积分中值定理得到. 因为 $f(u,v)$ 是连续的, 所以 (6) 近似为 $f(x,y)\Delta x\Delta y$. 事实上,

$$\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{1}{\Delta x\Delta y}|\Pr\{x \leqslant X \leqslant x+\Delta x, y \leqslant Y \leqslant y+\Delta y\}-f(x,y)\Delta x\Delta y|=0. \qquad (7)$$

现在我们考虑有 p 个变量 $X_1, X_2, \cdots, X_p$ 的情形. 其 cdf

$$F(x_1,\cdots,x_p)=\Pr\{X_1\leqslant x_1,\cdots,X_p\leqslant x_p\} \qquad (8)$$

定义在实数 $x_1,\cdots,x_p$ 的每个集合上. 如果 $F(x_1,\cdots,x_p)$ 是绝对连续的, 其密度函数为

$$\frac{\partial^p F(x_1,\cdots,x_p)}{\partial x_1\cdots\partial x_p}=f(x_1,\cdots,x_p) \qquad (9)$$

(几乎处处), 从而

$$F(x_1,\cdots,x_p)=\int_{-\infty}^{x_p}\cdots\int_{-\infty}^{x_1} f(u_1,\cdots,u_p)\mathrm{d}u_1\cdots\mathrm{d}u_p. \qquad (10)$$

落在 p 维欧氏空间中任意 (可测) 集合 R 上的概率是

$$\Pr\{(X_1,\cdots,X_p)\in R\}=\int\underset{R}{\cdots}\int f(x_1,\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p. \qquad (11)$$

如果 $f(x_1,\cdots,x_p)$ 是连续的, 则概率元素 $f(x_1,\cdots,x_p)\Delta x_1\cdots\Delta x_p$ 近似为概率 $\Pr\{x_1\leqslant X_1\leqslant x_1+\Delta x_1,\cdots,x_p\leqslant X_p\leqslant x_p+\Delta x_p\}$. 联合矩定义为①

$$\mathrm{E}(X_1^{h_1}\cdots X_p^{h_p})=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{h_1}\cdots x_p^{h_p} f(x_1,\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p. \qquad (12)$$

① E 表示数学期望.

2.2.2 边缘分布

给定两个随机变量 X, Y 的 cdf 为 $F(x,y)$, 则 X 的边缘 cdf 为

$$\begin{aligned}\Pr\{X \leqslant x\} &= \Pr\{X \leqslant x, Y \leqslant \infty\} \\ &= F(x, \infty).\end{aligned} \tag{13}$$

记为 $F(x)$. 显然

$$F(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, v)\mathrm{d}v\mathrm{d}u. \tag{14}$$

我们称

$$\int_{-\infty}^{\infty} f(u, v)\mathrm{d}v = f(u) \tag{15}$$

为 X 的*边缘密度*. 则 (14) 为

$$F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u. \tag{16}$$

我们可用同样的方法定义 Y 的边缘 cdf $G(y)$ 和边缘密度 $g(y)$.

现在我们考虑广义情形. 给定 $X_1, \cdots, X_p$ 的 cdf $F(x_1, \cdots, x_p)$, 我们希望求 $X_1, \cdots, X_p$ 中某些变量, 比如说 $X_1, \cdots, X_r (r < p)$ 的边缘 cdf. 它是

$$\begin{aligned}&\Pr\{X_1 \leqslant x_1, \cdots, X_r \leqslant x_r\} \\ &\quad = \Pr\{X_1 \leqslant x_1, \cdots, X_r \leqslant x_r, X_{r+1} \leqslant \infty, \cdots, X_p \leqslant \infty\} \\ &\quad = F(x_1, \cdots, x_r, \infty, \cdots, \infty).\end{aligned} \tag{17}$$

$X_1, \cdots, X_r$ 的边缘密度是

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_r, u_{r+1}, \cdots, u_p)\mathrm{d}u_{r+1} \cdots \mathrm{d}u_p. \tag{18}$$

$X_1, \cdots, X_p$ 的任意其他子集的边缘分布和边缘密度可用同样方法得到.

变量子集的联合矩可由边缘分布求得, 例如,

$$\begin{aligned}\mathrm{E}(X_1^{h_1} \cdots X_r^{h_r}) &= \mathrm{E}(X_1^{h_1} \cdots X_r^{h_r} X_{r+1}^0 \cdots X_p^0) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{h_1} \cdots x_r^{h_r} f(x_1, \cdots, x_p)\mathrm{d}x_1 \cdots \mathrm{d}x_p \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{h_1} \cdots x_r^{h_r} \\ &\quad \cdot \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_p)\mathrm{d}x_{r+1} \cdots \mathrm{d}x_p\right] \mathrm{d}x_1 \cdots \mathrm{d}x_r.\end{aligned} \tag{19}$$

2.2.3 统计独立性

cdf 为 $F(x,y)$ 的两个随机变量 X, Y 称为*独立的*, 如果

$$F(x, y) = F(x)G(y), \tag{20}$$

其中 $F(x)$ 是 X 的边缘 cdf, $G(y)$ 是 Y 的边缘 cdf. 这意味着 X, Y 的密度是

$$\begin{aligned} f(x,y) &= \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 F(x)G(y)}{\partial x \partial y} \\ &= \frac{\mathrm{d}F(x)}{\mathrm{d}x} \cdot \frac{\mathrm{d}G(y)}{\mathrm{d}y} \\ &= f(x)g(y). \end{aligned} \tag{21}$$

相反地, 如果 $f(x,y) = f(x)g(y)$, 则

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\mathrm{d}u\mathrm{d}v = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u)g(v)\mathrm{d}u\mathrm{d}v \\ &= \int_{-\infty}^{x} f(u)\mathrm{d}u \int_{-\infty}^{y} g(v)\mathrm{d}v = F(x)G(y). \end{aligned} \tag{22}$$

因此, 当密度存在时, 独立性的一个等价定义就是 $f(x,y) = f(x)g(y)$. 为了更清楚统计独立性的深层意义, 对于任意给定的 $x_1 < x_2, y_1 < y_2$, 我们考虑概率

$$\begin{aligned} &\Pr\{x_1 \leqslant X \leqslant x_2, y_1 \leqslant Y \leqslant y_2\} \\ &= \int_{y_1}^{y_2}\int_{x_1}^{x_2} f(u,v)\mathrm{d}u\mathrm{d}v = \int_{x_1}^{x_2} f(u)\mathrm{d}u \int_{y_1}^{y_2} g(v)\mathrm{d}v \\ &= \Pr\{x_1 \leqslant X \leqslant x_2\}\Pr\{y_1 \leqslant Y \leqslant y_2\}. \end{aligned} \tag{23}$$

X 落在给定区间且 Y 落在另一个给定区间的概率, 就是 X 落在给定区间的概率与 Y 落在另一个给定区间的概率的乘积.

假设 $X_1, \cdots, X_p$ 的 cdf 是 $F(x_1, \cdots, x_p)$, 这个随机变量集合称为**相互独立的**, 如果

$$F(x_1, \cdots, x_p) = F_1(x_1) \cdots F_p(x_p), \tag{24}$$

其中 $F_i(x_i)$ 是 X_i 的边缘 cdf, $i = 1, \cdots, p$. 子集 $X_1, \cdots, X_r$ 称为与子集 $X_{r+1}, \cdots, X_p$ 独立的, 如果

$$F(x_1, \cdots, x_p) = F(x_1, \cdots, x_r, \infty, \cdots, \infty) \cdot F(\infty, \cdots, \infty, x_{r+1}, \cdots, x_p). \tag{25}$$

独立性的另一个结果是关于联合矩的. 例如, 如果 $X_1, \cdots, X_p$ 是相互独立的, 则

$$\begin{aligned} \mathrm{E}(X_1^{h_1} \cdots X_p^{h_p}) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{h_1} \cdots x_p^{h_p} f_1(x_1) \cdots f_p(x_p)\mathrm{d}x_1 \cdots \mathrm{d}x_p \\ &= \prod_{i=1}^{p} \int_{-\infty}^{\infty} x_i^{h_i} f_i(x_i)\mathrm{d}x_i \\ &= \prod_{i=1}^{p} \{\mathrm{E}(X_i^{h_i})\}. \end{aligned} \tag{26}$$

2.2.4　条件分布

如果 A 和 B 是两个事件, 其中 A 和 B 同时发生的概率是 $P(AB)$, B 发生的概率是 $P(B) > 0$, 则已知 B 发生时 A 发生的条件概率是 $P(AB)/P(B)$. 假设事

件 A 是 X 落在区间 $[x_1,x_2]$ 里, 事件 B 是 Y 落在区间 $[y_1,y_2]$ 里. 则 Y 落在区间 $[y_1,y_2]$ 时 X 落在区间 $[x_1,x_2]$ 的条件概率是

$$\begin{aligned}\Pr\{x_1 \leqslant X \leqslant x_2 | y_1 \leqslant Y \leqslant y_2\} &= \frac{\Pr\{x_1 \leqslant X \leqslant x_2, y_1 \leqslant Y \leqslant y_2\}}{\Pr\{y_1 \leqslant Y \leqslant y_2\}} \\ &= \frac{\int_{x_1}^{x_2}\int_{y_1}^{y_2} f(u,v)\mathrm{d}v\mathrm{d}u}{\int_{y_1}^{y_2} g(v)\mathrm{d}v}.\end{aligned} \tag{27}$$

现在设 $y_1 = y, y_2 = y + \Delta y$. 则对连续密度, 有

$$\int_y^{y+\Delta y} g(v)\mathrm{d}v = g(y^*)\Delta y, \tag{28}$$

其中 $y \leqslant y^* \leqslant y + \Delta y$. 同样

$$\int_y^{y+\Delta y} f(u,v)\mathrm{d}v = f[u, y^*(u)]\Delta y, \tag{29}$$

其中 $y \leqslant y^*(u) \leqslant y + \Delta y$. 因此,

$$\Pr\{x_1 \leqslant X \leqslant x_2 | y \leqslant Y \leqslant y + \Delta y\} = \int_{x_1}^{x_2} \frac{f[u, y^*(u)]}{g(y^*)}\mathrm{d}u. \tag{30}$$

注意到, 对于固定的 y 和 Δy (> 0), (30) 中的被积函数可视为一个单变量密度函数. 现在对于满足 $g(y) > 0$ 的 y, 我们定义 $\Pr\{x_1 \leqslant X \leqslant x_2 | Y = y\}$, 即给定 Y 等于 y 时 X 落在 x_1 和 x_2 之间的概率, 为 $\Delta y \to 0$ 时 (30) 的极限. 因此

$$\Pr\{x_1 \leqslant X \leqslant x_2 | Y = y\} = \int_{x_1}^{x_2} f(u|y)\mathrm{d}u, \tag{31}$$

其中 $f(u|y) = f(u,y)/g(y)$. 对于给定的 y, $f(u|y)$ 是一个密度函数, 称它为给定 y 时 X 的**条件密度**. 注意到, 若 X 和 Y 相互独立, 则 $f(x|y) = f(x)$.

在一般情况下, 若随机变量 $X_1, \cdots, X_p$ 的 cdf 是 $F(x_1, \cdots, x_p)$, 则给定 $X_{r+1} = x_{r+1}, \cdots, X_p = x_p$ 时, $X_1, \cdots, X_r$ 的条件密度是

$$\frac{f(x_1, \cdots, x_p)}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(u_1, \cdots, u_r, x_{r+1}, \cdots, x_p)\mathrm{d}u_1\cdots\mathrm{d}u_r}. \tag{32}$$

关于条件概率的更多讨论, 读者可以参考 Chung (1974), Kolmogorov (1950), Loève (1977), Loève(1978) 和 Neveu (1965).

2.2.5 变量变换

设 $X_1, \cdots, X_p$ 的密度为 $f(x_1, \cdots, x_p)$. 考虑 p 元实值函数

$$y_i = y_i(x_1, \cdots, x_p), \quad i = 1, \cdots, p. \tag{33}$$

假定从 x 空间到 y 空间的变换是一一对应的,[①] 其逆变换为

$$x_i = x_i(y_1, \cdots, y_p), \quad i = 1, \cdots, p. \tag{34}$$

将随机变量 $Y_1, \cdots, Y_p$ 定义为

① 更准确地说, 假定这在 x 空间中 $f(x_1, \cdots, x_p)$ 为正的那部分上成立的.

$$Y_i = y_i(X_1, \cdots, X_p), \quad i = 1, \cdots, p. \tag{35}$$

则 $Y_1, \cdots, Y_p$ 的密度是

$$g(y_1, \cdots, y_p) = f[x_1(y_1, \cdots, y_p), \cdots, x_p(y_1, \cdots, y_p)]J(y_1, \cdots, y_p), \tag{36}$$

其中 $J(y_1, \cdots, y_p)$ 是雅可比行列式

$$J(y_1, \cdots, y_p) = \mathrm{mod}\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \cdots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}. \tag{37}$$

我们假定这些导数存在, 这里 mod 运算表示对它后面的表达式取模或绝对值. $(X_1, \cdots, X_p)$ 落在一个区域 R 中的概率由 (11) 给出, 则 $(Y_1, \cdots, Y_p)$ 落在区域 S 的概率是

$$\Pr\{(Y_1, \cdots, Y_p) \in S\} = \int \underset{S}{\cdots} \int g(y_1, \cdots, y_p)\mathrm{d}y_1 \cdots \mathrm{d}y_p. \tag{38}$$

若 S 是 R 的变换, 即 R 中每一点经过变换 (33) 都变为 S 中的一点, 且 S 中的每一点经过变换 (34) 都变成 R 中的一点, 则由普通的多重积分变换理论可知 (11) 等于 (38). 这也说明了 (36) 确实是 $Y_1, \cdots, Y_p$ 的密度.

2.3　多元正态分布

单变量正态密度函数可以写成形式

$$k\mathrm{e}^{-\frac{1}{2}\alpha(x-\beta)^2} = k\mathrm{e}^{-\frac{1}{2}(x-\beta)\alpha(x-\beta)}, \tag{1}$$

其中 α 是一个正数, k 由 (1) 在整个 x 轴上积分等于 1 所决定. $X_1, \cdots, X_p$ 的多元正态分布的密度函数有一个类推形式. 标量变量 x 被向量

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \tag{2}$$

替代, 标量常数 β 被向量

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix} \tag{3}$$

替代, 正常数 α 被一个正定 (对称) 矩阵

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix} \tag{4}$$

替代. 平方项 $\alpha(x-\beta)^2=(x-\beta)\alpha(x-\beta)$ 被二次型

$$(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})=\sum_{i,j=1}^{p}a_{ij}(x_i-b_i)(x_j-b_j) \tag{5}$$

替代. 从而 p 维正态分布的密度函数是

$$f(x_1,\cdots,x_p)=K\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})}, \tag{6}$$

其中 $K(>0)$ 是由上式在整个 $x_1,\cdots,x_p$ 的 p 维欧氏空间上积分等于 1 所决定的.

从矩阵形式的表达式中可以清楚地看到多元正态分布的密度 (6) 和单变量正态分布的密度 (1) 的相似性. 矩阵符号和运算将会贯穿本书. 读者可以参考附录中矩阵理论的回顾和矩阵运算中符号的定义.

我们注意到 $f(x_1,\cdots,x_p)$ 是非负的. 由于 $\boldsymbol{A}$ 是正定的, 则有

$$(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})\geqslant 0, \tag{7}$$

因此密度 (6) 是有界的, 即

$$f(x_1,\cdots,x_p)\leqslant K. \tag{8}$$

现在我们来计算使得 (6) 在整个 p 维空间的积分等于 1 的 K 值. 首先计算

$$K^*=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})}\mathrm{d}x_p\cdots\mathrm{d}x_1. \tag{9}$$

我们用到了下面事实 (见附录中的推论 A.1.6), 即如果 $\boldsymbol{A}$ 是正定的, 则存在一个非奇异矩阵 $\boldsymbol{C}$, 使得

$$\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}=\boldsymbol{I}, \tag{10}$$

其中 $\boldsymbol{I}$ 表示单位矩阵, $\boldsymbol{C}'$ 表示 $\boldsymbol{C}$ 的转置. 令

$$\boldsymbol{x}-\boldsymbol{b}=\boldsymbol{C}\boldsymbol{y}, \tag{11}$$

其中

$$\boldsymbol{y}=\begin{pmatrix} y_1\\ \vdots\\ y_p\end{pmatrix}. \tag{12}$$

则

$$(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})=\boldsymbol{y}'\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}\boldsymbol{y}=\boldsymbol{y}'\boldsymbol{y}. \tag{13}$$

这个变换的雅可比行列式是

$$\boldsymbol{J}=\operatorname{mod}|\boldsymbol{C}|, \tag{14}$$

其中 $\operatorname{mod}|\boldsymbol{C}|$ 表示 $\boldsymbol{C}$ 的行列式的绝对值. 因此 (9) 变为

$$K^*=\operatorname{mod}|\boldsymbol{C}|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}}\mathrm{d}y_p\cdots\mathrm{d}y_1. \tag{15}$$

我们有

$$\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}}=\exp\left(-\frac{1}{2}\sum_{i=1}^{p}y_i^2\right)=\prod_{i=1}^{p}\mathrm{e}^{-\frac{1}{2}y_i^2}, \tag{16}$$

其中 $\exp(z)=\mathrm{e}^z$. 我们将 (15) 写为

$$K^* = \text{mod}|\boldsymbol{C}| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_1^2} \cdots e^{-\frac{1}{2}y_p^2} dy_p \cdots dy_1 \tag{17}$$
$$= \text{mod}|\boldsymbol{C}| \prod_{i=1}^{p} \left\{ \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i \right\}$$
$$= \text{mod}|\boldsymbol{C}| \prod_{i=1}^{p} \{\sqrt{2\pi}\}$$
$$= \text{mod}|\boldsymbol{C}|(2\pi)^{\frac{1}{2}p}$$

这里利用了

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 1. \tag{18}$$

由 (10) 可得行列式方程

$$|\boldsymbol{C}'| \cdot |\boldsymbol{A}| \cdot |\boldsymbol{C}| = |\boldsymbol{I}|. \tag{19}$$

因为

$$|\boldsymbol{C}'| = |\boldsymbol{C}|, \tag{20}$$

$|\boldsymbol{I}| = 1$, 可由 (19) 得到

$$\text{mod}|\boldsymbol{C}| = 1/\sqrt{|\boldsymbol{A}|}. \tag{21}$$

从而

$$K = 1/K^* = \sqrt{|\boldsymbol{A}|}(2\pi)^{-\frac{1}{2}p}. \tag{22}$$

则正态密度函数为

$$\frac{\sqrt{|\boldsymbol{A}|}}{(2\pi)^{\frac{1}{2}p}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{b})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{b})}. \tag{23}$$

我们将通过求 $X_1, \cdots, X_p$ 的一阶矩和二阶矩来说明 $\boldsymbol{b}$ 和 $\boldsymbol{A}$ 的重要性. 为了方便, 可将这些随机变量当作一个随机向量

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}. \tag{24}$$

更一般地, 我们可以定义随机矩阵和随机矩阵的期望值, 随机向量可以看作列数为 1 的特殊随机矩阵.

定义 2.3.1 一个随机矩阵 $\boldsymbol{Z}$ 是一个关于随机变量 $Z_{11}, \cdots, Z_{mn}$ 的矩阵,

$$\boldsymbol{Z} = (Z_{gh}), \qquad g = 1, \cdots, m, \quad h = 1, \cdots, n. \tag{25}$$

如果随机变量 $Z_{11}, \cdots, Z_{mn}$ 只取有限个值, 则随机矩阵 $\boldsymbol{Z}$ 的取值范围是有限个矩阵, 比方说这些矩阵为 $\boldsymbol{Z}(1), \cdots, \boldsymbol{Z}(q)$. 如果 $\boldsymbol{Z} = \boldsymbol{Z}(i)$ 的概率是 p_i, 则我们可以定义 $\text{E}(\boldsymbol{Z})$ 为 $\sum_{i=1}^{q} \boldsymbol{Z}(i)p_i$. 则 $\text{E}(\boldsymbol{Z}) = (\text{E}(Z_{gh}))$. 如果随机变量 $Z_{11}, \cdots, Z_{mn}$ 有联合密度, 则通过黎曼和运算, 可以定义 $\text{E}(\boldsymbol{Z})$ 为离散情况下渐近和的极限 (如果此极限存在的话), 同样有 $\text{E}(\boldsymbol{Z}) = (\text{E}(Z_{gh}))$. 从而我们可以用下面的一般定义.

定义 2.3.2 随机矩阵 $\boldsymbol{Z}$ 的期望值是

$$\mathrm{E}(\boldsymbol{Z})=(\mathrm{E}(Z_{gh})), \qquad g=1,\cdots,m, \quad h=1,\cdots,n. \tag{26}$$

特别地, 如果 $\boldsymbol{Z}$ 是由 (24) 定义的 $\boldsymbol{X}$, 则称其期望值

$$\mathrm{E}(\boldsymbol{X})=\begin{pmatrix}\mathrm{E}(X_1)\\ \vdots \\ \mathrm{E}(X_p)\end{pmatrix} \tag{27}$$

为 $\boldsymbol{X}$ 的**均值**或**均值向量**. 我们通常用 $\boldsymbol{\mu}$ 来表示这个均值向量. 如果 $\boldsymbol{Z}$ 是 $(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'$, 则其期望值是

$$\mathrm{C}(\boldsymbol{X})=\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'=[\mathrm{E}(X_i-\mu_i)(X_j-\mu_j)], \tag{28}$$

称之为 $\boldsymbol{X}$ 的**协方差**或**协方差阵**. 这个矩阵的第 i 个对角线元素 $\mathrm{E}(X_i-\mu_i)^2$ 是 X_i 的**方差**; 而第 i,j 个非对角线元素 $\mathrm{E}(X_i-\mu_i)(X_j-\mu_j)$ 是 X_i 和 X_j 的**协方差**, $i\neq j$. 我们常把这个协方差阵表示为 $\boldsymbol{\Sigma}$. 注意到

$$\mathrm{C}(\boldsymbol{X})=\mathrm{E}(\boldsymbol{X}\boldsymbol{X}'-\boldsymbol{\mu}\boldsymbol{X}'-\boldsymbol{X}\boldsymbol{\mu}'+\boldsymbol{\mu}\boldsymbol{\mu}')=\mathrm{E}(\boldsymbol{X}\boldsymbol{X}')-\boldsymbol{\mu}\boldsymbol{\mu}'. \tag{29}$$

随机矩阵 (向量) 的期望值的运算满足某些法则, 我们将其列在下面的引理中.

引理 2.3.1 *如果 $\boldsymbol{Z}$ 是一个 $m\times n$ 随机矩阵, $\boldsymbol{D}$ 是一个 $l\times m$ 实矩阵, $\boldsymbol{E}$ 是一个 $n\times q$ 实矩阵, $\boldsymbol{F}$ 是一个 $l\times q$ 实矩阵, 则*

$$\mathrm{E}(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{E}+\boldsymbol{F})=\boldsymbol{D}(\mathrm{E}(\boldsymbol{Z}))\boldsymbol{E}+\boldsymbol{F}. \tag{30}$$

证明 $\mathrm{E}(\boldsymbol{D}\boldsymbol{Z}\boldsymbol{E}+\boldsymbol{F})$ 的第 i 行第 j 列元素是

$$\mathrm{E}\left(\sum_{h,g} d_{ih}Z_{hg}e_{gj}+f_{ij}\right)=\sum_{h,g} d_{ih}(\mathrm{E}(Z_{hg}))e_{gj}+f_{ij}, \tag{31}$$

它就是 $\boldsymbol{D}(\mathrm{E}(\boldsymbol{Z}))\boldsymbol{E}+\boldsymbol{F}$ 的第 i 行第 j 列元素. ■

引理 2.3.2 *如果 $\boldsymbol{Y}=\boldsymbol{D}\boldsymbol{X}+\boldsymbol{f}$, 其中 $\boldsymbol{X}$ 是一个随机向量, 则*

$$\mathrm{E}(\boldsymbol{Y})=\boldsymbol{D}\mathrm{E}(\boldsymbol{X})+\boldsymbol{f}, \tag{32}$$

$$\mathrm{C}(\boldsymbol{Y})=\boldsymbol{D}\mathrm{C}(\boldsymbol{X})\boldsymbol{D}'. \tag{33}$$

证明 第一个结论可以直接由引理 2.3.1 得出, 至于第二个结论, 由于

$$\begin{aligned}\mathrm{C}(\boldsymbol{Y})&=\mathrm{E}(\boldsymbol{Y}-\mathrm{E}(\boldsymbol{Y}))(\boldsymbol{Y}-\mathrm{E}(\boldsymbol{Y}))' \\ &=\mathrm{E}[\boldsymbol{D}\boldsymbol{X}+\boldsymbol{f}-(\boldsymbol{D}\mathrm{E}(\boldsymbol{X})+\boldsymbol{f})][\boldsymbol{D}\boldsymbol{X}+\boldsymbol{f}-(\boldsymbol{D}\mathrm{E}(\boldsymbol{X})+\boldsymbol{f})]' \\ &=\mathrm{E}[\boldsymbol{D}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))][\boldsymbol{D}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))]' \\ &=\mathrm{E}[\boldsymbol{D}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))'\boldsymbol{D}'],\end{aligned} \tag{34}$$

由引理 2.3.1 可知结论 (33) 成立. ■

做变换 (11), 即 $\boldsymbol{X}=\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{b}$, 则有 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{C}\mathrm{E}(\boldsymbol{Y})+\boldsymbol{b}$. 由 2.2 节的变换理论可知, $\boldsymbol{Y}$ 的密度与 (16) 成比例, 即为

$$\frac{1}{(2\pi)^{\frac{1}{2}p}}\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}}=\prod_{j=1}^{p}\left\{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}y_j^2}\right\}. \tag{35}$$

$\boldsymbol{Y}$ 的第 i 个分量的期望值是

$$\begin{aligned}\mathrm{E}(Y_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i \prod_{j=1}^{p} \left\{ \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y_j^2} \right\} \mathrm{d}y_1 \cdots \mathrm{d}y_p \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \prod_{\substack{j=1\\ j\neq i}}^{p} \left\{ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y_j^2} \mathrm{d}y_j \right\} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \\ &= 0.\end{aligned} \tag{36}$$

最后一个等式成立是因为[①] $y_i\mathrm{e}^{-\frac{1}{2}y_i^2}$ 是 y_i 的奇函数. 因此 $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{0}$. 从而 $\boldsymbol{X}$ 的均值 $\boldsymbol{\mu}$ 是

$$\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}) = \boldsymbol{b}. \tag{37}$$

从 (33) 中, 我们知道 $\mathcal{C}(\boldsymbol{X}) = \boldsymbol{C}(\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}'))\boldsymbol{C}'$. 因为 $\boldsymbol{Y}$ 的密度是 (35), 所以 $\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')$ 的第 i,j 个元素是

$$\mathrm{E}(Y_iY_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i y_j \prod_{h=1}^{p} \left\{ \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y_h^2} \right\} \mathrm{d}y_1 \cdots \mathrm{d}y_p. \tag{38}$$

如果 $i = j$, 我们有

$$\begin{aligned}\mathrm{E}(Y_i^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \prod_{\substack{h=1\\ h\neq i}}^{p} \left\{ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y_h^2} \mathrm{d}y_h \right\} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \\ &= 1.\end{aligned} \tag{39}$$

最后一个等式成立是因为前一个表达式是一个均值为 0 且方差为 1 的正态分布变量的平方. 如果 $i \neq j$, 则 (38) 变为

$$\begin{aligned}\mathrm{E}(Y_iY_j) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \mathrm{e}^{-\frac{1}{2}y_i^2} \mathrm{d}y_i \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_j \mathrm{e}^{-\frac{1}{2}y_j^2} \mathrm{d}y_j \\ &\quad \cdot \prod_{\substack{h=1\\ h\neq i,j}}^{p} \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}y_h^2} \mathrm{d}y_h \right\} \\ &= 0, \quad i \neq j,\end{aligned} \tag{40}$$

这是因为第一个积分为 0. 我们可以把 (39) 和 (40) 概括为

$$\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') = \boldsymbol{I}. \tag{41}$$

因此

$$\mathrm{E}(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})' = \boldsymbol{C}\boldsymbol{I}\boldsymbol{C}' = \boldsymbol{C}\boldsymbol{C}'. \tag{42}$$

① 或者说, 最后一个等式成立是因为其前一表达式是一个均值为 0 的正态分布变量的期望值.

将等式 (10) 的左右两边同时左乘 $(\boldsymbol{C}')^{-1}$, 右乘 $\boldsymbol{C}^{-1}$, 我们得到 $\boldsymbol{A}=(\boldsymbol{C}')^{-1}\boldsymbol{C}^{-1}$. 等式的左右两边同时取逆, 我们得到

$$\boldsymbol{C}\boldsymbol{C}'=\boldsymbol{A}^{-1}. \tag{43}$$

则 $\boldsymbol{X}$ 的协方差阵为

$$\boldsymbol{\Sigma}=\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'=\boldsymbol{A}^{-1}. \tag{44}$$

从 (43) 中, 我们可以看出 $\boldsymbol{\Sigma}$ 是正定的. 让我们总结一下这些结果.

定理 2.3.1 *如果一个 p 维随机向量 $\boldsymbol{X}$ 的密度是 (23), 则 $\boldsymbol{X}$ 的期望值是 $\boldsymbol{b}$, 协方差阵是 $\boldsymbol{A}^{-1}$. 反之, 给定一个向量 $\boldsymbol{\mu}$ 和一个正定矩阵 $\boldsymbol{\Sigma}$, 存在一个多元正态密度*

$$(2\pi)^{-\frac{1}{2}p}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \tag{45}$$

使得具有这个密度的随机向量的期望值是 $\boldsymbol{\mu}$ 且协方差阵是 $\boldsymbol{\Sigma}$.

我们将把密度 (45) 表示为 $n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ 并且把其分布律表示为 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

协方差阵的第 i 个对角元 σ_{ii} 是 $\boldsymbol{X}$ 的第 i 个分量的方差, 我们有时可以把它表示为 σ_i^2. X_i 和 X_j 的相关系数定义为

$$\rho_{ij}=\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}=\frac{\sigma_{ij}}{\sigma_i\sigma_j}. \tag{46}$$

这个关联性测度关于 X_i 和 X_j 是对称的: $\rho_{ij}=\rho_{ji}$. 因为

$$\begin{pmatrix}\sigma_{ii} & \sigma_{ij}\\ \sigma_{ji} & \sigma_{jj}\end{pmatrix}=\begin{pmatrix}\sigma_i^2 & \sigma_i\sigma_j\rho_{ij}\\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2\end{pmatrix} \tag{47}$$

是正定的 (附录的推论 A.1.3), 所以行列式

$$\begin{vmatrix}\sigma_i^2 & \sigma_i\sigma_j\rho_{ij}\\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2\end{vmatrix}=\sigma_i^2\sigma_j^2(1-\rho_{ij}^2) \tag{48}$$

是正的. 因此, $-1<\rho_{ij}<1$. (关于奇异分布, 见 2.4 节.) 多元正态密度可以由均值 μ_i $(i=1,\cdots,p)$、方差 σ_i^2 $(i=1,\cdots,p)$ 和相关系数 ρ_{ij} $(i<j,\ i,j=1,\cdots,p)$ 参数化.

作为前面理论的特殊情况, 我们考虑二元正态分布. 均值向量为

$$\mathrm{E}\begin{pmatrix}X_1\\ X_2\end{pmatrix}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}; \tag{49}$$

协方差阵可以写为

$$\begin{aligned}\boldsymbol{\Sigma}&=\mathrm{E}\begin{pmatrix}(X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2)\\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2\end{pmatrix}\\ &=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix},\end{aligned} \tag{50}$$

其中 σ_1^2 是 X_1 的方差, σ_2^2 是 X_2 的方差, ρ 是 X_1 和 X_2 的相关系数. (50) 的逆是

$$\mathbf{\Sigma}^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix}\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2}\\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2}\end{pmatrix}. \tag{51}$$

X_1 和 X_2 的密度函数是

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \tag{52}$$

定理 2.3.2 任何二元分布的相关系数 ρ 关于变换 $X_i^*=b_iX_i+c_i(b_i>0,i=1,2)$ 是不变的. 对于一个二元正态分布, 如果其参数的函数, 关于前述变换是不变的, 则该函数一定是 ρ 的某个函数.

证明 由引理 2.3.2 可知, X_i^* 的方差是 $b_i^2\sigma_i^2, i=1,2$, X_1^* 和 X_2^* 的协方差是 $b_1b_2\sigma_1\sigma_2\rho$. 将这些值代入 X_1^* 和 X_2^* 的相关系数的定义中, 得到它们的相关系数就是 ρ. 如果 $f(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ 关于上述变换不变, 则取 $b_i=1/\sigma_i$ 和 $c_i=-\mu_i/\sigma_i, i=1,2$, 其一定变为 $f(0,0,1,1,\rho)$. ■

相关系数 ρ 是 X_1 和 X_2 之间自然的**关联性测度**. 任何二元正态分布的参数函数, 如果它独立于散度和位置参数, 则它是 ρ 的一个函数. **标准化变量** (标准得分) 是 $Y_i=(X_i-\mu_i)/\sigma_i$. 任意两个标准化变量的均方差为

$$\mathrm{E}(Y_1-Y_2)^2=2(1-\rho). \tag{53}$$

(53) 越小 (即 ρ 越大), Y_1 和 Y_2 就越相似. 如果 $\rho>0$, X_1 和 X_2 倾向于正相关; 如果 $\rho<0$, X_1 和 X_2 倾向于负相关. 如果 $\rho=0$, 密度 (52) 是 X_1 和 X_2 边缘密度的乘积, 所以 X_1 和 X_2 独立.

注意到对于任何一个正值 c, 密度函数 (45) 在 p 维欧氏空间里的椭球面

$$(\boldsymbol{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=c \tag{54}$$

上是常数. 每个椭球面的中心都是点 $\boldsymbol{\mu}$. 椭球面的形状和定向由 $\mathbf{\Sigma}$ 决定, 大小 (给定 $\mathbf{\Sigma}$ 时) 由 c 决定. 如果 $\mathbf{\Sigma}=\sigma^2\boldsymbol{I}$, 则 (54) 是一个球面, 从而 $n(\boldsymbol{x}|\boldsymbol{\mu},\sigma^2\boldsymbol{I})$ 称为**球面正态密度**.

让我们来详细考虑密度 (52) 的二元情形. 我们变换坐标 $(x_i-\mu_i)/\sigma_i=y_i$, $i=1,2$, 以便常数密度轨迹的中心在原点. 这些轨迹可以由

$$\frac{1}{1-\rho^2}(y_1^2-2\rho y_1y_2+y_2^2)=c \tag{55}$$

决定. 其在 y_1 轴和 y_2 轴上的截距相等. 如果 $\rho>0$, 椭圆长轴沿 45° 方向, 长度为 $2\sqrt{c(1+\rho)}$, 短轴长度为 $2\sqrt{c(1-\rho)}$. 如果 $\rho<0$, 长轴沿 135° 方向, 长度为 $2\sqrt{c(1-\rho)}$, 短轴长度为 $2\sqrt{c(1+\rho)}$. ρ 的大小决定了这些长度的比值. 在二元情形里, 我们可以认为密度函数是平面上的曲面. 其密度等高线是地形图上同海拔等高线, 它们说明了山 (概率曲面) 的形状. 如果 $\rho>0$, 山的走向是一条正斜率直线, 大多数山是在第一象限和第三象限. 当我们变换回来 $x_i=\sigma_iy_i+\mu_i$, 我们将每一个等高线沿第 i 个轴的方向以因子 σ_i 伸长, 并将中心平移至 (μ_1,μ_2).

单正态变量的 cdf 的数值可在大多数统计教材的表中找到. 数值

$$F(x_1,x_2)=\Pr\{X_1\leqslant x_1,X_2\leqslant x_2\} \quad (56)$$
$$=\Pr\left\{\frac{X_1-\mu_1}{\sigma_1}\leqslant y_1,\frac{X_2-\mu_2}{\sigma_2}\leqslant y_2\right\},$$

其中 $y_1=(x_1-\mu_1)/\sigma_1$ 和 $y_2=(x_2-\mu_2)/\sigma_2$, 可以在 Pearson (1931) 中找到. (美国) 国家标准局 (1959) 给出了一个广泛的表. Gupta (1963) 给出了一个关于这些表的参考书目. Pearson 还指出了

$$F(x_1,x_2)=\sum_{j=0}^{\infty}\rho^j\tau_j(y_1)\tau_j(y_2), \quad (57)$$

其中所谓的**四项函数**$\tau_j(y)$ 在 Pearson(1930) 中列至 $\tau_{19}(y)$. Harris and Soms(1980) 研究了 (57) 的推广.

2.4 正态分布变量线性组合的分布, 变量的独立性, 边缘分布

研究正态多元分布是很有意义的, 其中一个原因是由多元正态分布得到的边缘分布和条件分布也是正态分布. 而且, 多元正态变量的线性组合也是正态分布的. 首先, 我们来说明, 如果对一个分量联合分布具有正态密度的向量做非奇异线性变换, 将得到一个分量联合分布也具有正态密度的向量.

定理 2.4.1 设 $\boldsymbol{X}$(有 p 个分量) 的分布是 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. 则对于非奇异阵 $\boldsymbol{C}$,

$$\boldsymbol{Y}=\boldsymbol{C}\boldsymbol{X} \quad (1)$$

的分布是 $N(\boldsymbol{C}\boldsymbol{\mu},\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$.

证明 把 $\boldsymbol{X}$ 的密度 $n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ 作变换,

$$\boldsymbol{x}=\boldsymbol{C}^{-1}\boldsymbol{y}, \quad (2)$$

再乘以变换 (2) 的雅可比行列式,

$$\operatorname{mod}|\boldsymbol{C}^{-1}|=\frac{1}{\operatorname{mod}|\boldsymbol{C}|}=\sqrt{\frac{1}{|\boldsymbol{C}|^2}}=\sqrt{\frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{C}|\cdot|\boldsymbol{\Sigma}|\cdot|\boldsymbol{C}'|}}=\frac{|\boldsymbol{\Sigma}|^{\frac12}}{|\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'|^{\frac12}}, \quad (3)$$

就得到 $\boldsymbol{Y}$ 的密度. $n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的指数中的二次型是

$$Q=(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}). \quad (4)$$

变换 (2) 将 Q 变为

$$\begin{aligned}Q&=(\boldsymbol{C}^{-1}\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}^{-1}\boldsymbol{y}-\boldsymbol{\mu}) \quad (5)\\&=(\boldsymbol{C}^{-1}\boldsymbol{y}-\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}^{-1}\boldsymbol{y}-\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{\mu})\\&=[\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})]\\&=(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})'(\boldsymbol{C}^{-1})'\boldsymbol{\Sigma}^{-1}\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})\\&=(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})'(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-1}(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu}),\end{aligned}$$

这是因为对 $\boldsymbol{C}\boldsymbol{C}^{-1}=\boldsymbol{I}$ 两边同时转置可得 $(\boldsymbol{C}^{-1})'=(\boldsymbol{C}')^{-1}$. 因此, $\boldsymbol{Y}$ 的密度是

$$\begin{aligned} & n(\boldsymbol{C}^{-1}\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})\operatorname{mod}|\boldsymbol{C}|^{-1} \qquad (6) \\ &= (2\pi)^{-\frac{1}{2}p}|\boldsymbol{C\Sigma C}'|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{C\mu})'(\boldsymbol{C\Sigma C}')^{-1}(\boldsymbol{y}-\boldsymbol{C\mu})\right] \\ &= n(\boldsymbol{y}|\boldsymbol{C\mu},\boldsymbol{C\Sigma C}'). \end{aligned}$$

■

现在我们考虑两个向量形式的随机变量集合 $X_1,\cdots,X_q$ 和 $X_{q+1},\cdots,X_p$,

$$\boldsymbol{X}^{(1)}=\begin{pmatrix} X_1 \\ \vdots \\ X_q \end{pmatrix},\quad \boldsymbol{X}^{(2)}=\begin{pmatrix} X_{q+1} \\ \vdots \\ X_p \end{pmatrix}. \qquad (7)$$

这些变量组成一个随机向量

$$\boldsymbol{X}=\begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix}=\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}. \qquad (8)$$

现在我们假定这 p 个变量服从联合正态分布, 其均值向量为

$$\mathrm{E}(\boldsymbol{X}^{(1)})=\boldsymbol{\mu}^{(1)},\quad \mathrm{E}(\boldsymbol{X}^{(2)})=\boldsymbol{\mu}^{(2)}, \qquad (9)$$

协方差阵为

$$\mathrm{E}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})'=\boldsymbol{\Sigma}_{11}, \qquad (10)$$

$$\mathrm{E}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})'=\boldsymbol{\Sigma}_{22}, \qquad (11)$$

$$\mathrm{E}(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})'=\boldsymbol{\Sigma}_{12}. \qquad (12)$$

我们称随机向量 $\boldsymbol{X}$ 被划分为如 (8) 的子向量, 类似地,

$$\boldsymbol{\mu}=\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \qquad (13)$$

被划分为子向量,

$$\boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \qquad (14)$$

被划分为子矩阵. 这里 $\boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}'$. (见附录, A.3 节.)

我们将说明, 如果 $\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}'=\boldsymbol{0}$, 则 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 是独立正态分布的. 于是

$$\boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \qquad (15)$$

它的逆是

$$\boldsymbol{\Sigma}^{-1}=\begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix}. \qquad (16)$$

因此 $n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的指数中的二次型是

$$\begin{aligned} Q &= (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\ &= \left[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})', (\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\right] \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)} \\ \boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)} \end{pmatrix} \\ &= \left[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}_{11}^{-1}, (\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_{22}^{-1}\right] \begin{pmatrix} \boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)} \\ \boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)} \end{pmatrix} \\ &= (\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)}) + (\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}) \\ &= Q_1 + Q_2, \end{aligned} \tag{17}$$

其中

$$\begin{aligned} Q_1 &= (\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)}), \\ Q_2 &= (\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}). \end{aligned} \tag{18}$$

我们还注意到 $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}|\cdot|\boldsymbol{\Sigma}_{22}|$. 则 $\boldsymbol{X}$ 的密度可以写为

$$\begin{aligned} n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) &= \frac{1}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}Q} \\ &= \frac{1}{(2\pi)^{\frac{1}{2}q}|\boldsymbol{\Sigma}_{11}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}Q_1}\cdot\frac{1}{(2\pi)^{\frac{1}{2}(p-q)}|\boldsymbol{\Sigma}_{22}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}Q_2} \\ &= n(\boldsymbol{x}^{(1)}|\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{11})n(\boldsymbol{x}^{(2)}|\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22}). \end{aligned} \tag{19}$$

$\boldsymbol{X}^{(1)}$ 的边缘密度由积分

$$\begin{aligned} &\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})\mathrm{d}x_{q+1}\cdots\mathrm{d}x_p \\ &\quad= n(\boldsymbol{x}^{(1)}|\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{11})\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} n(\boldsymbol{x}^{(2)}|\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22})\mathrm{d}x_{q+1}\cdots\mathrm{d}x_p \\ &\quad= n(\boldsymbol{x}^{(1)}|\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{11}) \end{aligned} \tag{20}$$

给出. 因此, $\boldsymbol{X}^{(1)}$ 的边缘分布是 $N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_{11})$, $\boldsymbol{X}^{(2)}$ 的边缘分布是 $N(\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22})$. $X_1,\cdots,X_p$ 的联合密度就是 $X_1,\cdots,X_q$ 的边缘密度和 $X_{q+1},\cdots,X_p$ 的边缘密度的乘积, 所以这两个变量集是独立的. 既然变量标号是任意的, 则 $\boldsymbol{X}^{(1)}$ 可以是任何子向量, 那么我们已经证明了下面定理的充分性.

定理 2.4.2 如果 $X_1,\cdots,X_p$ 服从联合正态分布, 则这些随机变量的一个子集和由剩余变量组成的子集相互独立的充要条件是这个子集的每一个变量和另一个子集的每一个变量的协方差为 0.

必要性依据如下事实, 如果 X_i 来自某一个子集, X_j 来自另一个子集, 则对任何密度 (见 2.2.3 节), 有

$$\begin{aligned} \sigma_{ij} &= \mathrm{E}(X_i-\mu_i)(X_j-\mu_j) \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_i-\mu_i)(x_j-\mu_j)f(x_1,\cdots,x_q) \\ &\quad\cdot f(x_{q+1},\cdots,x_p)\mathrm{d}x_1\cdots\mathrm{d}x_p \end{aligned} \tag{21}$$

$$
\begin{aligned}
&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_i-\mu_i)f(x_1,\cdots,x_q)\mathrm{d}x_1\cdots\mathrm{d}x_q\\
&\quad\cdot\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x_j-\mu_j)f(x_{q+1},\cdots,x_p)\mathrm{d}x_{q+1}\cdots\mathrm{d}x_p\\
&=0.
\end{aligned}
$$

因为 $\sigma_{ij}=\sigma_i\sigma_j\rho_{ij}$, $\sigma_i\neq 0,\sigma_j\neq 0$ (我们默认 $\boldsymbol{\Sigma}$ 是非奇异的), 所以条件 $\sigma_{ij}=0$ 等价于 $\rho_{ij}=0$. 因此, 如果一个变量集和剩余的变量不相关, 则这两个集合是独立的. 需要强调的是, 不相关就一定独立的结论是依赖正态性假设的, 但反之总是成立的.

我们来考虑二元正态分布的特殊情况. 这时 $\boldsymbol{X}^{(1)}=X_1$, $\boldsymbol{X}^{(2)}=X_2$, $\boldsymbol{\mu}^{(1)}=\mu_1$, $\boldsymbol{\mu}^{(2)}=\mu_2$, $\boldsymbol{\Sigma}_{11}=\sigma_{11}=\sigma_1^2$, $\boldsymbol{\Sigma}_{22}=\sigma_{22}=\sigma_2^2$ 和 $\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}=\sigma_{12}=\sigma_1\sigma_2\rho_{12}$. 因此, 如果 X_1 和 X_2 服从二元正态分布, 则它们相互独立当且仅当它们不相关. 如果它们不相关, X_i 的边缘分布是均值为 μ_i 和方差为 σ_i^2 的正态分布. 以上讨论也证明了下面的推论.

推论 2.4.1 *如果 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 且 $\boldsymbol{X}$ 的一组分量和剩余分量不相关, 则这组变量的边缘分布是多元正态分布, 其均值、方差和协方差分别是 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 中相对应的元素.*

现在我们来说明这个推论即使在两个变量集不独立的时候也成立. 我们像以前那样划分 $\boldsymbol{X}$, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$. 我们对子向量做非奇异线性变换

$$\boldsymbol{Y}^{(1)}=\boldsymbol{X}^{(1)}+\boldsymbol{B}\boldsymbol{X}^{(2)}, \tag{22}$$

$$\boldsymbol{Y}^{(2)}=\boldsymbol{X}^{(2)}, \tag{23}$$

选择 $\boldsymbol{B}$ 使得 $\boldsymbol{Y}^{(1)}$ 的分量与 $\boldsymbol{Y}^{(2)}=\boldsymbol{X}^{(2)}$ 的分量不相关. 则矩阵 $\boldsymbol{B}$ 必须满足等式

$$
\begin{aligned}
\mathbf{0}&=\mathrm{E}(\boldsymbol{Y}^{(1)}-\mathrm{E}(\boldsymbol{Y}^{(1)}))(\boldsymbol{Y}^{(2)}-\mathrm{E}(\boldsymbol{Y}^{(2)}))' \qquad (24)\\
&=\mathrm{E}(\boldsymbol{X}^{(1)}+\boldsymbol{B}\boldsymbol{X}^{(2)}-\mathrm{E}(\boldsymbol{X}^{(1)})-\boldsymbol{B}\mathrm{E}(\boldsymbol{X}^{(2)}))(\boldsymbol{X}^{(2)}-\mathrm{E}(\boldsymbol{X}^{(2)}))'\\
&=\mathrm{E}\left[(\boldsymbol{X}^{(1)}-\mathrm{E}(\boldsymbol{X}^{(1)}))+\boldsymbol{B}(\boldsymbol{X}^{(2)}-\mathrm{E}(\boldsymbol{X}^{(2)}))\right](\boldsymbol{X}^{(2)}-\mathrm{E}(\boldsymbol{X}^{(2)}))'\\
&=\boldsymbol{\Sigma}_{12}+\boldsymbol{B}\boldsymbol{\Sigma}_{22}.
\end{aligned}
$$

因此 $\boldsymbol{B}=-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ 且

$$\boldsymbol{Y}^{(1)}=\boldsymbol{X}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)}. \tag{25}$$

向量

$$\begin{pmatrix}\boldsymbol{Y}^{(1)}\\ \boldsymbol{Y}^{(2)}\end{pmatrix}=\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \mathbf{0} & \boldsymbol{I}\end{pmatrix}\boldsymbol{X} \tag{26}$$

是 $\boldsymbol{X}$ 的非奇异变换, 因此它服从正态分布, 且有

$$\mathrm{E}\begin{pmatrix}\boldsymbol{Y}^{(1)}\\ \boldsymbol{Y}^{(2)}\end{pmatrix}=\mathrm{E}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \mathbf{0} & \boldsymbol{I}\end{pmatrix}\boldsymbol{X} \tag{27}$$

$$=\begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}$$
$$=\begin{pmatrix} \boldsymbol{\mu}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}=\begin{pmatrix} \boldsymbol{\nu}^{(1)} \\ \boldsymbol{\nu}^{(2)} \end{pmatrix}$$
$$=\boldsymbol{\nu},$$

以及

$$\begin{aligned} \mathrm{C}(\boldsymbol{Y}) &= \mathrm{E}(\boldsymbol{Y}-\boldsymbol{\nu})(\boldsymbol{Y}-\boldsymbol{\nu})' \\ &= \begin{pmatrix} \mathrm{E}(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})' & \mathrm{E}(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})(\boldsymbol{Y}^{(2)}-\boldsymbol{\nu}^{(2)})' \\ \mathrm{E}(\boldsymbol{Y}^{(2)}-\boldsymbol{\nu}^{(2)})(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})' & \mathrm{E}(\boldsymbol{Y}^{(2)}-\boldsymbol{\nu}^{(2)})(\boldsymbol{Y}^{(2)}-\boldsymbol{\nu}^{(2)})' \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \end{aligned} \tag{28}$$

上式成立是因为

$$\begin{aligned} &\mathrm{E}(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})(\boldsymbol{Y}^{(1)}-\boldsymbol{\nu}^{(1)})' \\ &\quad= \mathrm{E}[(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})] \\ &\qquad\cdot[(\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})]' \\ &\quad= \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ &\quad= \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned} \tag{29}$$

因此, $\boldsymbol{Y}^{(1)}$ 和 $\boldsymbol{Y}^{(2)}$ 独立, 且由推论 2.4.1 知, $\boldsymbol{X}^{(2)}=\boldsymbol{Y}^{(2)}$ 有边缘分布 $N(\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22})$. 因为 $\boldsymbol{X}$ 的分量个数是任意的, 所以我们有下面定理.

定理 2.4.3 *如果 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 则 $\boldsymbol{X}$ 的任意分量的子集的边缘分布都是多元正态的, 其均值、方差和协方差分别为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 中相对应的元素.*

现在考虑变换

$$\boldsymbol{Z}=\boldsymbol{D}\boldsymbol{X}, \tag{30}$$

其中 $\boldsymbol{Z}$ 有 q 个分量, $\boldsymbol{D}$ 是 $q\times p$ 实矩阵. 则 $\boldsymbol{Z}$ 的期望值是

$$\mathrm{E}(\boldsymbol{Z})=\boldsymbol{D}\boldsymbol{\mu}, \tag{31}$$

协方差阵是

$$\mathrm{E}(\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{\mu})(\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{\mu})'=\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}'. \tag{32}$$

上面的方法也可用于 $q=p$ 且 $\boldsymbol{D}$ 是非奇异矩阵的情况. 如果 $q\leqslant p$ 且 $\boldsymbol{D}$ 的秩是 q, 我们可以找到一个 $(p-q)\times p$ 的矩阵 $\boldsymbol{E}$, 使得

$$\begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{W} \end{pmatrix}=\begin{pmatrix} \boldsymbol{D} \\ \boldsymbol{E} \end{pmatrix}\boldsymbol{X} \tag{33}$$

为非奇异矩阵. (见附录, A.3 节.) 则 $\boldsymbol{Z}$ 和 $\boldsymbol{W}$ 服从联合正态分布, 且由定理 2.4.3 知 $\boldsymbol{Z}$ 有边缘正态分布. 因此, 对于秩为 q 的 $\boldsymbol{D}$ (且 $\boldsymbol{X}$ 服从非奇异分布, 即有密度), 我们已证明了下面定理.

定理 2.4.4 如果 $\boldsymbol{X}$ 服从 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 分布，则 $\boldsymbol{Z}=\boldsymbol{D}\boldsymbol{X}$ 服从分布 $N(\boldsymbol{D}\boldsymbol{\mu},\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}')$，其中 $\boldsymbol{D}$ 是秩为 $q\leqslant p$ 的 $q\times p$ 矩阵.

下面讨论奇异或退化正态分布，并将定理 2.4.4 推广到 $\boldsymbol{D}$ 为任意矩阵的情形. 一个奇异分布是 p- 空间上集中在一个低维集合上的分布，即在任何与这个给定集合不相交集合上的概率都是 0. 对于奇异正态分布的情况，质量集中在一个给定的线性集合上 [即许多 $(p-1)$ 维超平面的交集]. 令 $\boldsymbol{y}$ 为这个线性集合里的坐标集(坐标的个数等于线性集合的维数)，则这个线性集合的参数定义可为 $\boldsymbol{x}=\boldsymbol{A}\boldsymbol{y}+\boldsymbol{\lambda}$，其中 $\boldsymbol{A}$ 是一个 $p\times q$ 矩阵，$\boldsymbol{\lambda}$ 是一个 p- 向量. 假定 $\boldsymbol{Y}$ 在 q 维线性集合上正态分布，则我们说

$$\boldsymbol{X}=\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{\lambda} \tag{34}$$

在 p 维空间里服从奇异或退化的正态分布. 如果 $\mathrm{E}(\boldsymbol{Y})=\boldsymbol{\nu}$，则有 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{A}\boldsymbol{\nu}+\boldsymbol{\lambda}=\boldsymbol{\mu}$. 如果 $\mathrm{E}(\boldsymbol{Y}-\boldsymbol{\nu})(\boldsymbol{Y}-\boldsymbol{\nu})'=\boldsymbol{T}$，则

$$\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'=\mathrm{E}[\boldsymbol{A}(\boldsymbol{Y}-\boldsymbol{\nu})(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{A}']=\boldsymbol{A}\boldsymbol{T}\boldsymbol{A}'=\boldsymbol{\Sigma}. \tag{35}$$

注意，如果 $p>q$，则 $\boldsymbol{\Sigma}$ 是奇异的且因此不可逆，故我们不能写出 $\boldsymbol{X}$ 的正态密度. 事实上，$\boldsymbol{X}$ 根本没有密度，因为其在任何与 q- 集合不相交集合上的概率都是 0，这说明其密度几乎处处为 0.

现在反过来看，如果 $\boldsymbol{X}$ 有均值 $\boldsymbol{\mu}$ 和秩为 r 的协方差阵 $\boldsymbol{\Sigma}$，它可以写成 (34) 的形式 (除 0 概率集外)，其中 $\boldsymbol{X}$ 是任意分布的. 如果 $\boldsymbol{\Sigma}$ 的秩是 r，则存在一个非奇异矩阵 $\boldsymbol{B}$ 使得

$$\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}'=\begin{pmatrix}\boldsymbol{I}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}, \tag{36}$$

其中单位矩阵的秩是 r. (见附录中的定理 A.4.1.) 变换

$$\boldsymbol{B}\boldsymbol{X}=\boldsymbol{V}=\begin{pmatrix}\boldsymbol{V}^{(1)}\\ \boldsymbol{V}^{(2)}\end{pmatrix} \tag{37}$$

定义了一个随机向量 $\boldsymbol{V}$，其协方差阵是 (36)，均值向量是

$$\mathrm{E}(\boldsymbol{V})=\boldsymbol{B}\boldsymbol{\mu}=\boldsymbol{\nu}=\begin{pmatrix}\boldsymbol{\nu}^{(1)}\\ \boldsymbol{\nu}^{(2)}\end{pmatrix}. \tag{38}$$

因为 $\boldsymbol{V}^{(2)}$ 中的协方差元素都是零，所以 $\boldsymbol{V}^{(2)}=\boldsymbol{\nu}^{(2)}$ 以概率 1 成立. 现在划分

$$\boldsymbol{B}^{-1}=(\boldsymbol{C}\ \ \boldsymbol{D}), \tag{39}$$

其中 $\boldsymbol{C}$ 有 r 列. 则 (37) 等价于

$$\boldsymbol{X}=\boldsymbol{B}^{-1}\boldsymbol{V}=(\boldsymbol{C}\ \ \boldsymbol{D})\begin{pmatrix}\boldsymbol{V}^{(1)}\\ \boldsymbol{V}^{(2)}\end{pmatrix}=\boldsymbol{C}\boldsymbol{V}^{(1)}+\boldsymbol{D}\boldsymbol{V}^{(2)}. \tag{40}$$

因此以概率 1 有

$$\boldsymbol{X}=\boldsymbol{C}\boldsymbol{V}^{(1)}+\boldsymbol{D}\boldsymbol{\nu}^{(2)}, \tag{41}$$

这里 $\boldsymbol{C}$ 就是形式 (34) 中的 $\boldsymbol{A}$, $\boldsymbol{V}^{(1)}$ 就是 $\boldsymbol{Y}$, $\boldsymbol{D}\boldsymbol{\nu}^{(2)}$ 就是 $\boldsymbol{\lambda}$.

现在我们给出一个涵盖奇异分布的正态分布的正式定义.

定义 2.4.1 随机向量 $\boldsymbol{X}$ 有 p 个分量, $\mathrm{E}(\boldsymbol{X})=\boldsymbol{\mu}$ 且 $\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'=\boldsymbol{\Sigma}$, 称它为正态分布的 [或为服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$], 如果存在一个变换 (34), 其中 $\boldsymbol{A}$ 的行数是 p, 列数是 $\boldsymbol{\Sigma}$ 的秩, 比如说 r, $\boldsymbol{Y}$ (有 r 个分量) 服从非奇异正态分布, 即有密度

$$k\mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\nu})'\boldsymbol{T}^{-1}(\boldsymbol{y}-\boldsymbol{\nu})}. \tag{42}$$

显然, 如果 $\boldsymbol{\Sigma}$ 的秩是 p, 则 $\boldsymbol{A}$ 可以取为 $\boldsymbol{I}$ 且 $\boldsymbol{\lambda}$ 取 0, 于是 $\boldsymbol{X}=\boldsymbol{Y}$, 从而定义 2.4.1 与 2.3 节一致. 为了避免定义 2.4.1 的冗余, 我们可以取 $\boldsymbol{T}=\boldsymbol{I}$ 和 $\boldsymbol{\nu}=\boldsymbol{0}$.

定理 2.4.5 如果 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 则 $\boldsymbol{Z}=\boldsymbol{D}\boldsymbol{X}$ 服从分布 $N(\boldsymbol{D}\boldsymbol{\mu},\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}')$.

证明 这个定理包括了 $\boldsymbol{X}$ 为非奇异或奇异分布的情况和 $\boldsymbol{D}$ 为非奇异或秩小于 q 的情况. 既然 $\boldsymbol{X}$ 可以用 (34) 表示, 其中 $\boldsymbol{Y}$ 服从非奇异分布 $N(\boldsymbol{\nu},\boldsymbol{T})$, 那么

$$\boldsymbol{Z}=\boldsymbol{D}\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{D}\boldsymbol{\lambda}, \tag{43}$$

其中 $\boldsymbol{D}\boldsymbol{A}$ 是 $q\times r$ 的. 如果 $\boldsymbol{D}\boldsymbol{A}$ 的秩是 r, 则定理得证. 如果秩小于 r, 如 s, 则 $\boldsymbol{Z}$ 的协方差阵,

$$\boldsymbol{D}\boldsymbol{A}\boldsymbol{T}\boldsymbol{A}'\boldsymbol{D}'=\boldsymbol{E} \tag{44}$$

的秩为 s. 由附录中的定理 A.4.1 可知, 存在一个非奇异矩阵

$$\boldsymbol{F}=\begin{pmatrix}\boldsymbol{F}_1\\ \boldsymbol{F}_2\end{pmatrix} \tag{45}$$

使得

$$\begin{aligned}\boldsymbol{F}\boldsymbol{E}\boldsymbol{F}'&=\begin{pmatrix}\boldsymbol{F}_1\boldsymbol{E}\boldsymbol{F}_1' & \boldsymbol{F}_1\boldsymbol{E}\boldsymbol{F}_2'\\ \boldsymbol{F}_2\boldsymbol{E}\boldsymbol{F}_1' & \boldsymbol{F}_2\boldsymbol{E}\boldsymbol{F}_2'\end{pmatrix}\\ &=\begin{pmatrix}(\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A})\boldsymbol{T}(\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A})' & (\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A})\boldsymbol{T}(\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})'\\ (\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})\boldsymbol{T}(\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A})' & (\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})\boldsymbol{T}(\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})'\end{pmatrix}=\begin{pmatrix}\boldsymbol{I}_s & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}.\end{aligned} \tag{46}$$

因此 $\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A}$ 的秩是 s(由附录中定理 A.1.1 的逆定理), 且 $\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A}=\boldsymbol{0}$, 这是因为 $(\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})\boldsymbol{T}(\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A})'$ 的每一个对角线元素都是 $\boldsymbol{F}_2\boldsymbol{D}\boldsymbol{A}$ 中某一行关于正定矩阵 $\boldsymbol{T}$ 的二次型. 因此 $\boldsymbol{F}\boldsymbol{Z}$ 的协方差阵是 (46), 且有

$$\boldsymbol{F}\boldsymbol{Z}=\begin{pmatrix}\boldsymbol{F}_1\\ \boldsymbol{F}_2\end{pmatrix}\boldsymbol{D}\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{F}\boldsymbol{D}\boldsymbol{\lambda}=\begin{pmatrix}\boldsymbol{F}_1\boldsymbol{D}\boldsymbol{A}\boldsymbol{Y}\\ \boldsymbol{0}\end{pmatrix}+\boldsymbol{F}\boldsymbol{D}\boldsymbol{\lambda}=\begin{pmatrix}\boldsymbol{U}_1\\ \boldsymbol{0}\end{pmatrix}+\boldsymbol{F}\boldsymbol{D}\boldsymbol{\lambda}. \tag{47}$$

显然 $\boldsymbol{U}_1$ 有一个非奇异正态分布. 令 $\boldsymbol{F}^{-1}=(\boldsymbol{G}_1\ \ \boldsymbol{G}_2)$. 则有形如 (34) 的

$$\boldsymbol{Z}=\boldsymbol{G}_1\boldsymbol{U}_1+\boldsymbol{D}\boldsymbol{\lambda}. \tag{48}$$

■

本节的讨论可以用上节的几何解释来说明. $\boldsymbol{X}$ 的密度在 2.3 节中的椭球面 (54) 上是常数. 因为变换 (2) 是线性变换 (即坐标轴的一个改变), 所以 $\boldsymbol{Y}$ 的密度在椭球面

$$(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})'(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')^{-1}(\boldsymbol{y}-\boldsymbol{C}\boldsymbol{\mu})=k \tag{49}$$

上是常数. $\boldsymbol{X}^{(1)}$ 的边缘分布是 $\boldsymbol{X}$ 的分布的质量在前 q 个坐标轴生成的 q 维空间上的投影. 常数密度的曲面也是椭球面, 其质量在任何直线上的投影都是正态的.

2.5 条件分布和多重相关系数

2.5.1 条件分布

本节我们求证由联合正态分布得到的条件分布也是正态的. 因为均值仅线性地依赖固定变量且方差和协方差根本不依赖这些固定变量的值, 所以这些条件分布有特别的简单结构. 本节讨论的偏相关理论和多重相关理论最早源于 Karl Pearson(1896) 对三个变量的研究, 后来被 Yule(1897a, 1897b) 推广.

设 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$($\boldsymbol{\Sigma}$ 非奇异). 我们像前面那样将 $\boldsymbol{X}$ 划分为 q 元子向量和 $(p-q)$ 元子向量

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{pmatrix}. \tag{1}$$

这里我们将用 2.4 节讨论的代数方法. $\boldsymbol{Y}^{(1)}=\boldsymbol{X}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)}$ 和 $\boldsymbol{Y}^{(2)}=\boldsymbol{X}^{(2)}$ 的联合密度是

$$n(\boldsymbol{y}^{(1)}|\boldsymbol{\mu}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})n(\boldsymbol{y}^{(2)}|\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22}).$$

则 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的密度可以通过变换这个表达式中的 $\boldsymbol{y}^{(1)}$ 为 $\boldsymbol{x}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{y}^{(2)}$ 为 $\boldsymbol{x}^{(2)}$ 而得到 (这个变换的雅可比行列是 1), 从而 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的密度是

$$\begin{aligned}f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})=&\frac{1}{(2\pi)^{\frac{1}{2}q}\sqrt{|\boldsymbol{\Sigma}_{11\cdot 2}|}}\exp\Big\{-\frac{1}{2}[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})]'\\&\cdot\boldsymbol{\Sigma}_{11\cdot 2}^{-1}[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})]\Big\}\\&\cdot\frac{1}{(2\pi)^{\frac{1}{2}(p-q)}\sqrt{|\boldsymbol{\Sigma}_{22}|}}\exp[-\frac{1}{2}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})],\end{aligned} \tag{2}$$

其中

$$\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{3}$$

这个密度一定是 $n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$. 给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件密度是 (2) 除以 $\boldsymbol{X}^{(2)}$ 在点 $\boldsymbol{x}^{(2)}$ 上的边缘密度 $n(\boldsymbol{x}^{(2)}|\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22})$ 即 (2) 的第二项的商. 这个商是

$$\begin{aligned}f(\boldsymbol{x}^{(1)}|\boldsymbol{x}^{(2)})=&\frac{1}{(2\pi)^{\frac{1}{2}q}\sqrt{|\boldsymbol{\Sigma}_{11\cdot 2}|}}\exp\Big\{-\frac{1}{2}[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})]'\\&\cdot\boldsymbol{\Sigma}_{11\cdot 2}^{-1}[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})]\Big\}.\end{aligned} \tag{4}$$

可以将 $\boldsymbol{x}^{(2)}$ 理解为是由 $p-q$ 个数组成的. 密度 $f(\boldsymbol{x}^{(1)}|\boldsymbol{x}^{(2)})$ 是一个 q- 变量正态密度, 其均值为

$$\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)})=\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})=\boldsymbol{\nu}(\boldsymbol{x}^{(2)}), \tag{5}$$

其协方差阵为

$$\mathrm{E}\{[\boldsymbol{X}^{(1)}-\boldsymbol{\nu}(\boldsymbol{x}^{(2)})][\boldsymbol{X}^{(1)}-\boldsymbol{\nu}(\boldsymbol{x}^{(2)})]'|\boldsymbol{x}^{(2)}\}=\boldsymbol{\Sigma}_{11\cdot2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{6}$$

需要注意, 给定 $\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的均值就是 $\boldsymbol{x}^{(2)}$ 的一个简单线性函数, 并且给定 $\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的协方差阵根本不依赖 $\boldsymbol{x}^{(2)}$.

定义 2.5.1 *矩阵 $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ 是 $\boldsymbol{X}^{(1)}$ 在 $\boldsymbol{x}^{(2)}$ 上的回归系数阵.*

$\boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ 的第 i 行第 $(k-q)$ 列的元素常表示为

$$\beta_{ik\cdot q+1,\cdots,k-1,k+1,\cdots,p},\quad i=1,\cdots,q,\ k=q+1,\cdots,p. \tag{7}$$

向量 $\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})$ 称为*回归函数*.

令 $\sigma_{ij\cdot q+1,\cdots,p}$ 是 $\boldsymbol{\Sigma}_{11\cdot2}$ 第 i,j 元. 我们称它们为*偏协方差*, $\sigma_{ii\cdot q+1,\cdots,p}$ 为*偏方差*.

定义 2.5.2

$$\rho_{ij\cdot q+1,\cdots,p}=\frac{\sigma_{ij\cdot q+1,\cdots,p}}{\sqrt{\sigma_{ii\cdot q+1,\cdots,p}}\sqrt{\sigma_{jj\cdot q+1,\cdots,p}}},\quad i,j=1,\cdots,q, \tag{8}$$

是给定 $X_{q+1},\cdots,X_p$ 时 X_i 和 X_j 的偏相关系数.

$\boldsymbol{X}$ 的分量个数是任意的且 q 的选取也是任意的. 因此, 上面的内容适合于定义给定任意 $\boldsymbol{X}$ 的 $p-q$ 个分量时, 其余 q 个分量的条件分布. 在偏协方差和偏相关系数中, 条件变量用圆点后面的下标表示. 在回归系数中因变量用第一个下标表示, 相关条件变量用第二个下标表示, 其余的条件变量用圆点后面的下标表示. 进而, 这些符号说明了任意 q 个变量在其他 $r-q$ 个变量 $(q\leqslant r\leqslant p)$ 条件下的条件分布.

定理 2.5.1 *划分 $\boldsymbol{X}$ 为两个子向量 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$. 假设相应地, 均值 $\boldsymbol{\mu}$ 也划分为 $\boldsymbol{\mu}^{(1)}$ 和 $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{X}$ 的协方差阵 $\boldsymbol{\Sigma}$ 划分为 $\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{12},\boldsymbol{\Sigma}_{22}$, 它们分别是 $\boldsymbol{X}^{(1)}$ 的协方差阵、$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的协方差阵以及 $\boldsymbol{X}^{(2)}$ 的协方差阵. 如果 $\boldsymbol{X}$ 的分布是正态的, 那么给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件分布也是正态的, 其均值为 $\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})$, 协方差阵为 $\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.*

作为一个例子, 我们考虑二元正态分布在给定 $X_2=x_2$ 时 X_1 的条件分布. 在这种情况下, $\boldsymbol{\mu}^{(1)}=\mu_1,\boldsymbol{\mu}^{(2)}=\mu_2,\boldsymbol{\Sigma}_{11}=\sigma_1^2,\boldsymbol{\Sigma}_{12}=\sigma_1\sigma_2\rho$ 和 $\boldsymbol{\Sigma}_{22}=\sigma_2^2$. 因此, 1×1 的回归系数阵是 $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}=\sigma_1\rho/\sigma_2$, 1×1 的偏协方差阵是

$$\boldsymbol{\Sigma}_{11\cdot2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}=\sigma_1^2-\sigma_1^2\sigma_2^2\rho^2/\sigma_2^2=\sigma_1^2(1-\rho^2). \tag{9}$$

给定 x_2 时 X_1 的密度是 $n[x_1|\mu_1+(\sigma_1\rho/\sigma_2)(x_2-\mu_2),\sigma_1^2(1-\rho^2)]$. 当 ρ 为正值时, 这个条件分布的均值随着 x_2 增长而增长; 当 ρ 为负值时, 随着 x_2 增长而减少. 注意当 $\sigma_1=\sigma_2$ 时, 例如, x_1 的条件分布的均值关于 μ_1 的增长不如 x_2 关于 μ_2 的增长多. [Galton (1889) 发现, 父代身高大于平均值的子代, 其平均身高趋向小于其父代的身高, 他称这种效应为 "趋近平庸的回归".] $|\rho|$ 越大, 条件分布的方差越小, 即 x_2 给出的关于 x_1 信息越多. 这是考虑用 ρ 来度量 X_1 和 X_2 关联性的另一个原因.

这个理论有一个启发性的几何解释. 密度 $f(x_1,x_2)$ 可以看作 x_1x_2 平面上的曲面 $z=f(x_1,x_2)$. 这个曲面与平面 $x_2=c$ 相交, 可以得到在 x_1x_2 平面中的直线 $x_2=c$ 上的曲线 $z=f(x_1,c)$. 这条曲线的纵坐标等比于给定 $x_2=c$ 时 X_1 的条件密度, 即一个单变量正态分布的密度曲线的纵坐标. 更一般地, 也很方便在 p 维空间里考虑常密度椭球面. 此时, 这些常密度 $f(x_1,\cdots,x_q|c_{q+1},\cdots,c_p)$ 的曲面是常密度 $f(x_1,\cdots,x_p)$ 的曲面与超平面 $x_{q+1}=c_{q+1},\cdots,x_p=c_p$ 的交集, 他们也是椭球面.

此外, 这些思想方法可能需要将实际总体理想化为是正态分布的. 例如, 考虑父一子对的总体. 如果这个总体可视为齐次的, 则父代的身高与对应子代的身高是近似正态分布的 (在某个范围内). 在所有父代的身高是 (比如说)5 英尺 9 英寸 (测量精度) 时, 子代身高的条件分布将是一个近似单变量正态分布. 这个正态分布的均值不同于所有父代身高 (比如说) 为 5 英尺 4 英寸的子代身高的均值, 但是它们的方差是相同的.

我们也可以考虑三元组的观测值, 父亲身高、其长子身高和次子身高. 在给定父代身高为 5 英尺 9 英寸下, 两组孩子身高的条件分布是两个变量的正态分布, 长子身高和次子身高的相关性是一个偏相关系数. 若固定父代的身高为常数, 从而消去了来自父代的遗传效应, 但是, 可以预测这个偏相关系数是正的, 因为母亲的遗传效应和环境因素将促使兄弟的身高有相似的变化.

我们在上面强调过, 任何从正态分布得到的条件分布都是正态的, 其均值是已固定变量的一个线性函数, 协方差阵是常值. 在非正态分布情形下, 一个变量集在另一个变量集上的条件分布通常没有这些性质. 但是, 可以构造出使得一些条件分布具有这些性质的非正态分布. 这个性质可以通过取 $\boldsymbol{X}$ 的密度为乘积 $n[\boldsymbol{x}^{(1)}|\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}),\boldsymbol{\Sigma}_{11\cdot2}]f(\boldsymbol{x}^{(2)})$ 而得到, 其中 $f(\boldsymbol{x}^{(2)})$ 是任意密度.

2.5.2 多重相关系数

我们再次把 $\boldsymbol{X}$ 划分为 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$. 我们将研究 $\boldsymbol{\beta X}^{(2)}$ 的一些性质.

定义 2.5.3　*向量 $\boldsymbol{X}^{(1\cdot2)}=\boldsymbol{X}^{(1)}-\boldsymbol{\mu}^{(1)}-\boldsymbol{\beta}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})$ 是向量 $\boldsymbol{X}^{(1)}$ 与其在 $\boldsymbol{X}^{(2)}$ 上的回归的残差向量.*

定理 2.5.2　*$\boldsymbol{X}^{(1\cdot2)}$ 的分量与 $\boldsymbol{X}^{(2)}$ 的分量不相关.*

证明　向量 $\boldsymbol{X}^{(1\cdot2)}$ 就是 2.4 节 (25) 中的 $\boldsymbol{Y}^{(1)}-\mathrm{E}(\boldsymbol{Y}^{(1)})$.

记 $\boldsymbol{\sigma}'_{(i)}$ 为 $\boldsymbol{\Sigma}_{12}$ 的第 i 行, $\boldsymbol{\beta}'_{(i)}$ 为 $\boldsymbol{\beta}$ 的第 i 行 (即 $\boldsymbol{\beta}'_{(i)}=\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}$). 令 $\mathrm{Var}(Z)$ 为 Z 的方差. ■

定理 2.5.3　*对任意向量 $\boldsymbol{\alpha}$,*

$$\mathrm{Var}(X_i^{(1\cdot2)})\leqslant\mathrm{Var}(X_i-\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}). \tag{10}$$

证明　由定理 2.5.2,

$$\begin{aligned}
&\operatorname{Var}(X_i-\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}) \qquad (11)\\
&=\mathrm{E}[X_i-\mu_i-\boldsymbol{\alpha}'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})]^2\\
&=\mathrm{E}[X_i^{(1\cdot2)}-\mathrm{E}(\boldsymbol{X}_i^{(1\cdot2)})+(\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha})'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})]^2\\
&=\operatorname{Var}[X_i^{(1\cdot2)}]+(\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha})'\mathrm{E}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})'(\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha})\\
&=\operatorname{Var}(X_i^{(1\cdot2)})+(\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha})'\boldsymbol{\Sigma}_{22}(\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha}).
\end{aligned}$$

由于 $\boldsymbol{\Sigma}_{22}$ 是正定的, 所以 $\boldsymbol{\beta}_{(i)}-\boldsymbol{\alpha}$ 的二次型是非负的且在 $\boldsymbol{\alpha}=\boldsymbol{\beta}_{(i)}$ 达到极小值 0. ■

因为 $\mathrm{E}(\boldsymbol{X}^{(1\cdot2)})=\mathbf{0}$, 所以 $\operatorname{Var}(X_i^{(1\cdot2)})=\mathrm{E}(X_i^{(1\cdot2)})^2$. 因此 $\boldsymbol{\mu}_i+\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})$ 是 $\boldsymbol{X}_i$ 在所有形如 $\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}+c$ 的 $\boldsymbol{X}^{(2)}$ 的函数中的**最优线性预测**, 其均方误差为零.

定理 2.5.4 *对任意向量 $\boldsymbol{\alpha}$*

$$\operatorname{Corr}(X_i,\boldsymbol{\beta}'_{(i)}\boldsymbol{X}^{(2)})\geqslant\operatorname{Corr}(X_i,\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}). \qquad (12)$$

证明 因为两个变量的相关系数在其中一个或两个都乘以一个正常数时不变, 所以我们可以假定 $\mathrm{E}[\boldsymbol{\alpha}'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})]^2=\mathrm{E}[\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})]^2$. 则 (10) 的展开式是

$$\begin{aligned}
&\sigma_{ii}-2\mathrm{E}(X_i-\mu_i)\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})+\operatorname{Var}(\boldsymbol{\beta}'_{(i)}\boldsymbol{X}^{(2)}) \qquad (13)\\
&\leqslant\sigma_{ii}-2\mathrm{E}(X_i-\mu_i)\boldsymbol{\alpha}'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})+\operatorname{Var}(\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}).
\end{aligned}$$

于是得到

$$\frac{\mathrm{E}(X_i-\mu_i)\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})}{\sqrt{\sigma_{ii}\operatorname{Var}(\boldsymbol{\beta}'_{(i)}\boldsymbol{X}^{(2)})}}\geqslant\frac{\mathrm{E}(X_i-\mu_i)\boldsymbol{\alpha}'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})}{\sqrt{\sigma_{ii}\operatorname{Var}(\boldsymbol{\alpha}'\boldsymbol{X}^{(2)})}}. \quad ■ \qquad (14)$$

定义 2.5.4 *X_i 和线性组合 $\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}$ 之间的极大相关系数称为 X_i 和 $\boldsymbol{X}^{(2)}$ 的多重相关系数.*

这就是说,

$$\begin{aligned}
\bar{R}_{i\cdot q+1,\cdots,p}&=\frac{\mathrm{E}[\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(X_i-\mu_i)]}{\sqrt{\sigma_{ii}}\sqrt{\mathrm{E}[\boldsymbol{\beta}'_{(i)}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\beta}_{(i)}]}} \qquad (15)\\
&=\frac{\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(i)}}{\sqrt{\sigma_{ii}}\sqrt{\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(i)}}}=\frac{\sqrt{\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(i)}}}{\sqrt{\sigma_{ii}}}.
\end{aligned}$$

一个有用的公式是

$$1-\bar{R}^2_{i\cdot q+1,\cdots,p}=\frac{\sigma_{ii}-\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(i)}}{\sigma_{ii}}=\frac{|\boldsymbol{\Sigma}_i|}{\sigma_{ii}|\boldsymbol{\Sigma}_{22}|}, \qquad (16)$$

其中将附录中的定理 A.3.2 应用到

$$\boldsymbol{\Sigma}_i=\begin{pmatrix}\sigma_{ii} & \boldsymbol{\sigma}'_{(i)}\\ \boldsymbol{\sigma}_{(i)} & \boldsymbol{\Sigma}_{22}\end{pmatrix}. \qquad (17)$$

因为

$$\sigma_{ii\cdot q+1,\cdots,p}=\sigma_{ii}-\boldsymbol{\sigma}'_{(i)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(i)}, \tag{18}$$

所以

$$\sigma_{ii\cdot q+1,\cdots,p}=(1-\bar{R}^2_{i\cdot q+1,\cdots,p})\sigma_{ii}. \tag{19}$$

这同时证明了, $\boldsymbol{X}$ 的任意分量的偏方差不可能大于方差. 实际上, $\bar{R}_{i\cdot q+1,\cdots,p}$ 越大, 条件分布的方差减少的就越多. 这也是用多重相关系数来测量 X_i 和 $\boldsymbol{X}^{(2)}$ 之间关联性的另一个原因.

$\boldsymbol{\beta}'_{(i)}\boldsymbol{X}^{(2)}$ 是 X_i 的最优线性预测, 并且是 X_i 与 $\boldsymbol{X}^{(2)}$ 的线性组合中的极大相关系数, 这一点仅依赖协方差的结构, 没有考虑正态性. 即使 $\boldsymbol{X}$ 不是正态分布的, $\boldsymbol{X}^{(1)}$ 在 $\boldsymbol{X}^{(2)}$ 上的回归也可以定义为 $\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})$, 其中, 残差由定义 2.5.3 给出, 偏协方差和相关系数可以定义为 (3) 和 (8) 中残差的协方差和相关系数. 不必从条件分布的观点解释这些量. 在正态情形下, $\mu_i+\boldsymbol{\beta}'_{(i)}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})$ 是 X_i 在给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时的条件期望. 如果不考虑正态性, $X_i-\mathrm{E}(X_i|\boldsymbol{X}^{(2)})$ 与 $\boldsymbol{X}^{(2)}$ 的任何函数都不相关, $\mathrm{E}(X_i|\boldsymbol{X}^{(2)})$ 关于 $\boldsymbol{X}^{(2)}$ 的函数 $h(\boldsymbol{X}^{(2)})$ 极小化 $\mathrm{E}[X_i-h(\boldsymbol{X}^{(2)})]^2$, 并且 $\mathrm{E}(X_i|\boldsymbol{X}^{(2)})$ 极大化 X_i 与 $\boldsymbol{X}^{(2)}$ 的函数的相关系数. (见习题 2.48~2.51.)

2.5.3 偏相关系数的一些公式

我们现在考虑一些给定不同变量集合时的条件分布的关系. 这些关系非常有用, 因为借助于它们, 我们可以从一个条件参数集合计算出另一个条件参数集合. 一个非常特别的情况是

$$\rho_{12\cdot 3}=\frac{\rho_{12}-\rho_{13}\rho_{23}}{\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23}^2}}, \tag{20}$$

这是 $p=3$, $q=2$ 时的 (8). 我们现在可以找到一个此结果的推广. 这个推导冗长, 但为了完整, 我们在这里还是给出来了.

令

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\\ \boldsymbol{X}^{(3)}\end{pmatrix}, \tag{21}$$

其中 $\boldsymbol{X}^{(1)}$ 有 p_1 个分量, $\boldsymbol{X}^{(2)}$ 有 p_2 个分量, $\boldsymbol{X}^{(3)}$ 有 p_3 个分量. 假定我们已知在给定 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的条件分布, 那么我们怎样找到在给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件分布? 我们可以利用在给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件密度为

$$\begin{aligned}f(\boldsymbol{x}^{(1)}|\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})&=\frac{f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})}{f(\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})}\\&=\frac{f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})/f(\boldsymbol{x}^{(3)})}{f(\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})/f(\boldsymbol{x}^{(3)})}\\&=\frac{f(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(3)})}{f(\boldsymbol{x}^{(2)}|\boldsymbol{x}^{(3)})}.\end{aligned} \tag{22}$$

在正态情形下, 给定 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的协方差阵为

$$\mathrm{Cov}\left[\left(\begin{array}{c}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{array}\right)\middle|\boldsymbol{x}^{(3)}\right]=\left(\begin{array}{cc}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{array}\right)-\left(\begin{array}{c}\boldsymbol{\Sigma}_{13}\\ \boldsymbol{\Sigma}_{23}\end{array}\right)\boldsymbol{\Sigma}_{33}^{-1}\left(\begin{array}{cc}\boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32}\end{array}\right) \quad (23)$$

$$=\left(\begin{array}{cc}\boldsymbol{\Sigma}_{11\cdot 3} & \boldsymbol{\Sigma}_{12\cdot 3}\\ \boldsymbol{\Sigma}_{21\cdot 3} & \boldsymbol{\Sigma}_{22\cdot 3}\end{array}\right),$$

其中,

$$\boldsymbol{\Sigma}=\left(\begin{array}{ccc}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23}\\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33}\end{array}\right). \quad (24)$$

给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 和 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件协方差可从给定 $\boldsymbol{X}^{(3)}=\boldsymbol{x}^{(3)}$ 时 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的条件协方差得到, 即

$$\mathrm{Cov}[\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)}]=\boldsymbol{\Sigma}_{11\cdot 3}-\boldsymbol{\Sigma}_{12\cdot 3}(\boldsymbol{\Sigma}_{22\cdot 3})^{-1}\boldsymbol{\Sigma}_{21\cdot 3}. \quad (25)$$

这个结果也说明了可由 $\sigma_{ij\cdot p_1+p_2,\cdots,p}$ $(i,j=1,\cdots,p_1+p_2)$ 计算出 $\sigma_{ij\cdot p_1+1,\cdots,p}$ $(i,j=1,\cdots,p_1)$.

特别地, 对 $p_1=q$, $p_2=1$ 和 $p_3=p-q-1$, 有

$$\sigma_{ij\cdot q+1,\cdots,p}=\sigma_{ij\cdot q+2,\cdots,p}-\frac{\sigma_{i,q+1\cdot q+2,\cdots,p}\sigma_{j,q+1\cdot q+2,\cdots,p}}{\sigma_{q+1,q+1\cdot q+2,\cdots,p}}, \quad (26)$$

$$i,j=1,\cdots,q.$$

因为

$$\sigma_{ii\cdot q+1,\cdots,p}=\sigma_{ii\cdot q+2,\cdots,p}(1-\rho^2_{i,q+1\cdot q+2,\cdots,p}), \quad (27)$$

所以

$$\rho_{ij\cdot q+1,\cdots,p}=\frac{\rho_{ij\cdot q+2,\cdots,p}-\rho_{i,q+1\cdot q+2,\cdots,p}\rho_{j,q+1\cdot q+2,\cdots,p}}{\sqrt{1-\rho^2_{i,q+1\cdot q+2,\cdots,p}}\sqrt{1-\rho^2_{j,q+1\cdot q+2,\cdots,p}}}. \quad (28)$$

这是一个有用的递推公式, 它可以由 $\{\rho_{ij}\}$ 连续地计算出 $\{\rho_{ij\cdot p}\},\{\rho_{ij\cdot p-1,p}\},\cdots$, $\rho_{12\cdot 3,\cdots,p}$.

2.6 特征函数和矩

2.6.1 特征函数

多元正态分布的特征函数与其密度函数有相似的形式. 由特征函数可以容易地得到矩和半不变量.

定义 2.6.1 对于任意实向量 $\boldsymbol{t}$, 随机向量 $\boldsymbol{X}$ 的特征函数是

$$\phi(\boldsymbol{t})=\mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}). \quad (1)$$

为使这个定义有意义, 我们先来定义随机向量的复值函数的期望值.

定义 2.6.2 将复值函数 $g(\boldsymbol{x})$ 写为 $g(\boldsymbol{x})=g_1(\boldsymbol{x})+\mathrm{i}g_2(\boldsymbol{x})$, 其中 $g_1(\boldsymbol{x})$ 和 $g_2(\boldsymbol{x})$ 是实值函数. 则 $g(\boldsymbol{X})$ 的期望值是

$$\mathrm{E}(g(\boldsymbol{X}))=\mathrm{E}(g_1(\boldsymbol{X}))+\mathrm{i}\,\mathrm{E}(g_2(\boldsymbol{X})). \tag{2}$$

特别地, 因为 $\mathrm{e}^{\mathrm{i}\theta}=\cos\theta+\mathrm{i}\sin\theta$, 则

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}})=\mathrm{E}(\cos\boldsymbol{t}'\boldsymbol{X})+\mathrm{i}\,\mathrm{E}(\sin\boldsymbol{t}'\boldsymbol{X}). \tag{3}$$

在计算向量 $\boldsymbol{X}$ 的特征函数时, 利用下面的引理通常很方便.

引理 2.6.1 设 $\boldsymbol{X}'=(\boldsymbol{X}^{(1)'}\boldsymbol{X}^{(2)'})$. 如果 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 互相独立且 $g(\boldsymbol{x})=g^{(1)}(\boldsymbol{x}^{(1)})g^{(2)}(\boldsymbol{x}^{(2)})$, 则

$$\mathrm{E}(g(\boldsymbol{X}))=\mathrm{E}(g^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g^{(2)}(\boldsymbol{X}^{(2)})). \tag{4}$$

证明 如果 $g(\boldsymbol{x})$ 是实值的且 $\boldsymbol{X}$ 有密度, 则

$$\begin{aligned}\mathrm{E}(g(\boldsymbol{X}))&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(\boldsymbol{x})f(\boldsymbol{x})\mathrm{d}x_1\cdots\mathrm{d}x_p \\ &=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g^{(1)}(\boldsymbol{x}^{(1)})g^{(2)}(\boldsymbol{x}^{(2)})f^{(1)}(\boldsymbol{x}^{(1)})f^{(2)}(\boldsymbol{x}^{(2)})\mathrm{d}x_1\cdots\mathrm{d}x_p \\ &=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g^{(1)}(\boldsymbol{x}^{(1)})f^{(1)}(\boldsymbol{x}^{(1)})\mathrm{d}x_1\cdots\mathrm{d}x_q \\ &\quad\cdot\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g^{(2)}(\boldsymbol{x}^{(2)})f^{(2)}(\boldsymbol{x}^{(2)})\mathrm{d}x_{q+1}\cdots\mathrm{d}x_p \\ &=\mathrm{E}(g^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g^{(2)}(\boldsymbol{X}^{(2)})).\end{aligned} \tag{5}$$

如果 $g(\boldsymbol{x})$ 是复值的,

$$\begin{aligned}g(\boldsymbol{x})&=[g_1^{(1)}(\boldsymbol{x}^{(1)})+\mathrm{i}g_2^{(1)}(\boldsymbol{x}^{(1)})][g_1^{(2)}(\boldsymbol{x}^{(2)})+\mathrm{i}g_2^{(2)}(\boldsymbol{x}^{(2)})] \\ &=g_1^{(1)}(\boldsymbol{x}^{(1)})g_1^{(2)}(\boldsymbol{x}^{(2)})-g_2^{(1)}(\boldsymbol{x}^{(1)})g_2^{(2)}(\boldsymbol{x}^{(2)}) \\ &\quad+\mathrm{i}[g_2^{(1)}(\boldsymbol{x}^{(1)})g_1^{(2)}(\boldsymbol{x}^{(2)})+g_1^{(1)}(\boldsymbol{x}^{(1)})g_2^{(2)}(\boldsymbol{x}^{(2)})].\end{aligned} \tag{6}$$

则

$$\begin{aligned}\mathrm{E}(g(\boldsymbol{X}))&=\mathrm{E}\left[g_1^{(1)}(\boldsymbol{X}^{(1)})g_1^{(2)}(\boldsymbol{X}^{(2)})-g_2^{(1)}(\boldsymbol{X}^{(1)})g_2^{(2)}(\boldsymbol{X}^{(2)})\right] \\ &\quad+\mathrm{i}\mathrm{E}\left[g_2^{(1)}(\boldsymbol{X}^{(1)})g_1^{(2)}(\boldsymbol{X}^{(2)})+g_1^{(1)}(\boldsymbol{X}^{(1)})g_2^{(2)}(\boldsymbol{X}^{(2)})\right]. \\ &=\mathrm{E}(g_1^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g_1^{(2)}(\boldsymbol{X}^{(2)}))-\mathrm{E}(g_2^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g_2^{(2)}(\boldsymbol{X}^{(2)})) \\ &\quad+\mathrm{i}\left[\mathrm{E}(g_2^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g_1^{(2)}(\boldsymbol{X}^{(2)}))+\mathrm{E}(g_1^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g_2^{(2)}(\boldsymbol{X}^{(2)}))\right]. \\ &=\left[\mathrm{E}(g_1^{(1)}(\boldsymbol{X}^{(1)}))+\mathrm{i}\mathrm{E}(g_2^{(1)}(\boldsymbol{X}^{(1)}))\right]\left[\mathrm{E}(g_1^{(2)}(\boldsymbol{X}^{(2)}))+\mathrm{i}\mathrm{E}(g_2^{(2)}(\boldsymbol{X}^{(2)}))\right] \\ &=\mathrm{E}(g^{(1)}(\boldsymbol{X}^{(1)}))\mathrm{E}(g^{(2)}(\boldsymbol{X}^{(2)})).\end{aligned} \tag{7}$$

■

将引理 2.6.1 连续地应用到 $g(\boldsymbol{X})=\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}$, 就得到下面的结论.

引理 2.6.2 如果 $\boldsymbol{X}$ 的分量是相互独立的, 则

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \prod_{j=1}^{p} \mathrm{E}(\mathrm{e}^{\mathrm{i}t_j' X_j}). \tag{8}$$

我们现在来求服从正态分布的随机向量的特征函数.

定理 2.6.1 若 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 则它的特征函数是

$$\phi(\boldsymbol{t}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}, \tag{9}$$

其中 $\boldsymbol{t}$ 是任意的实向量.

证明 我们从附录中的推论 A.1.6 可知, 存在非奇异矩阵 $\boldsymbol{C}$ 使得

$$\boldsymbol{C}'\boldsymbol{\Sigma}^{-1}\boldsymbol{C} = \boldsymbol{I}. \tag{10}$$

因此

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{C}'^{-1}\boldsymbol{C}^{-1} = (\boldsymbol{C}\boldsymbol{C}')^{-1}. \tag{11}$$

设

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{Y}, \tag{12}$$

则 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{0}, \boldsymbol{I})$.

现在 $\boldsymbol{Y}$ 的特征函数是

$$\psi(\boldsymbol{u}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{u}'\boldsymbol{Y}}) = \prod_{j=1}^{p} \mathrm{E}(\mathrm{e}^{\mathrm{i}u_j Y_j}). \tag{13}$$

因为 Y_j 服从分布 $N(0,1)$, 所以

$$\psi(\boldsymbol{u}) = \prod_{j=1}^{p} \mathrm{e}^{-\frac{1}{2}u_j^2} = \mathrm{e}^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}}. \tag{14}$$

于是对于 $\boldsymbol{t}'\boldsymbol{C} = \boldsymbol{u}'$,

$$\begin{aligned} \phi(\boldsymbol{t}) &= \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'(\boldsymbol{C}\boldsymbol{Y}+\boldsymbol{\mu})}) \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{C}\boldsymbol{Y}}) \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{t}'\boldsymbol{C})(\boldsymbol{t}'\boldsymbol{C})'}, \end{aligned} \tag{15}$$

第三个等号是通过把其两边都写成积分而得到的. 又由 (11) 有

$$\begin{aligned} \phi(\boldsymbol{t}) &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\mathrm{e}^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{C}\boldsymbol{C}'\boldsymbol{t}} \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}}. \end{aligned} \tag{16}$$

定理得证. ■

正态分布的特征函数非常有用. 例如, 我们可以用这种证明方法说明 2.4 节的结果. 如果 $\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{X}$, 则 $\boldsymbol{Z}$ 的特征函数是

$$\begin{aligned} \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}}) &= \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{D}\boldsymbol{X}}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}(\boldsymbol{D}'\boldsymbol{t})'\boldsymbol{X}}) \\ &= \mathrm{e}^{\mathrm{i}(\boldsymbol{D}'\boldsymbol{t})'\boldsymbol{\mu} - \frac{1}{2}(\boldsymbol{D}'\boldsymbol{t})'\boldsymbol{\Sigma}(\boldsymbol{D}'\boldsymbol{t})} \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}'(\boldsymbol{D}\boldsymbol{\mu}) - \frac{1}{2}\boldsymbol{t}'(\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}')\boldsymbol{t}}, \end{aligned} \tag{17}$$

这是 $N(\boldsymbol{D}\boldsymbol{\mu}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}')$ 的特征函数 (由定理 2.6.1).

用特征函数证明只有多元正态分布才具有每一个变量的线性组合也是正态分布的性质是很有意思的. 考虑一个 p 维向量 $\boldsymbol{Y}$, 其密度为 $f(\boldsymbol{y})$ 且特征函数为

$$\psi(\boldsymbol{u}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}\boldsymbol{u}'\boldsymbol{Y}}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\boldsymbol{u}'\boldsymbol{y}} f(\boldsymbol{y}) \mathrm{d}y_1 \cdots \mathrm{d}y_p, \tag{18}$$

并且假定 $\boldsymbol{y}$ 的均值是 $\boldsymbol{\mu}$, 协方差阵是 $\boldsymbol{\Sigma}$. 假设 $\boldsymbol{u}'\boldsymbol{Y}$ 对每个 $\boldsymbol{u}$ 都是正态分布的. 则这个线性组合的特征函数是

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}t\boldsymbol{u}'\boldsymbol{Y}}) = \mathrm{e}^{\mathrm{i}t\boldsymbol{u}'\boldsymbol{\mu} - \frac{1}{2}t^2\boldsymbol{u}'\boldsymbol{\Sigma}\boldsymbol{u}}. \tag{19}$$

现在令 $t=1$, 则等号右边是 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的特征函数, 结果得证 (由前面的定理 2.6.1 和后面的定理 2.6.3).

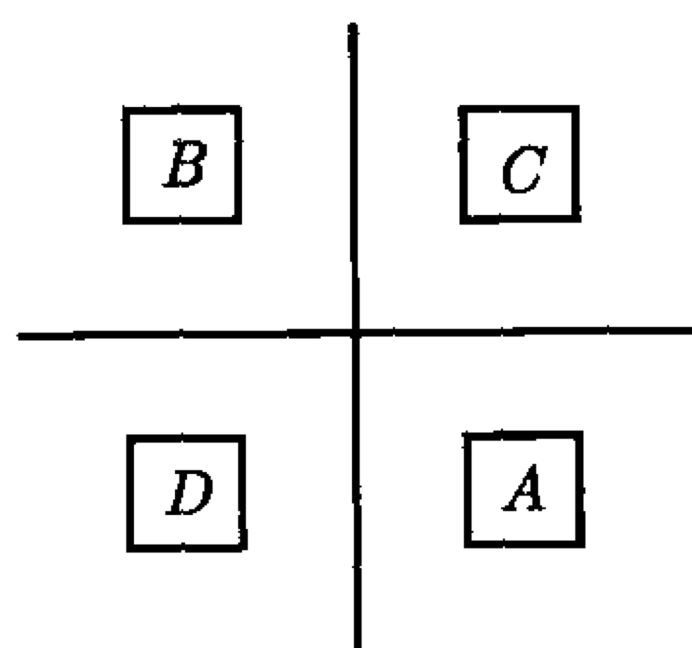


图 2.1

定理 2.6.2 *如果向量 $\boldsymbol{Y}$ 的每个分量线性组合都是正态分布的, 则 $\boldsymbol{Y}$ 也是正态分布的.*

需要强调的是定理 2.6.2 中, 对每个线性组合都是正态分布的要求是必须的. 例如, 如果 $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)'$ 且 $\boldsymbol{Y}_1$ 与 $\boldsymbol{Y}_2$ 不是相互独立的, 则 $\boldsymbol{Y}_1$ 和 $\boldsymbol{Y}_2$ 可以各有一个边缘正态分布. 有一个可以很容易用几何表示的例子. 设 $\boldsymbol{X}_1, \boldsymbol{X}_2$ 服从均值为零的联合正态分布. 在图 2.1 中, 从矩形 A 到 C, B 到 D 转移相同的质量. 可以看到最后 $\boldsymbol{Y}$ 的分布满足 $\boldsymbol{Y}_1$ 与 $\boldsymbol{Y}_2$ 的边缘分布分别和 $\boldsymbol{X}_1$ 与 $\boldsymbol{X}_2$ 一样, 都是正态的, 但是 $\boldsymbol{Y}_1$ 与 $\boldsymbol{Y}_2$ 的联合分布却不是正态的.

这个例子也可以用来说明两个变量, $\boldsymbol{Y}_1$ 与 $\boldsymbol{Y}_2$, 它们可以不相关且具有正态边缘分布. 但是它们的联合分布不一定是正态的, 它们也不一定是独立的. 这只需要选择合适的矩形, 使得对于得到的分布, $\boldsymbol{Y}_1\boldsymbol{Y}_2$ 的期望值为零. 从几何角度来看, 这是可以实现的.

为了以后参考, 我们给出两个关于特征函数的有用定理.

定理 2.6.3 *如果随机向量 $\boldsymbol{X}$ 有密度 $f(\boldsymbol{x})$ 且特征函数是 $\phi(\boldsymbol{t})$, 则*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\boldsymbol{t}'\boldsymbol{x}} \phi(\boldsymbol{t}) \mathrm{d}t_1 \cdots \mathrm{d}t_p. \tag{20}$$

这说明了特征函数唯一决定密度函数. 如果 $\boldsymbol{X}$ 没有密度函数, 则特征函数唯一决定其在任何连续区间上的概率. 在单变量情形下, 一个连续区间是一个端点不是 cdf 间断点的区间.

定理 2.6.4 *设 $\{F_j(\boldsymbol{x})\}$ 是一列 cdf , 且令 $\{\phi_j(\boldsymbol{t})\}$ 是一列相应的特征函数. $F_j(\boldsymbol{x})$ 收敛到一个 cdf $F(\boldsymbol{x})$ 的充分必要条件是, 对每个 $\boldsymbol{t}$, $\phi_j(\boldsymbol{t})$ 收敛到一个在 $\boldsymbol{t}=\boldsymbol{0}$ 连续的极限 $\phi(\boldsymbol{t})$ 上. 当这个条件满足时, 极限 $\phi(\boldsymbol{t})$ 就是极限分布 $F(\boldsymbol{x})$ 的特征函数.*

关于这两个定理的证明, 读者可以参考 Cramér(1946) 的 10.6 节和 10.7 节.

2.6.2 矩和半不变量

具有联合正态分布的 $X_1, \cdots, X_p$ 的矩可从特征函数 (9) 中得到. 其均值是

$$\mathrm{E}(X_h)=\frac{1}{\mathrm{i}}\frac{\partial\phi}{\partial t_h}\bigg|_{t=0} \tag{21}$$
$$=\frac{1}{\mathrm{i}}\left\{-\sum_j \sigma_{hj}t_j+\mathrm{i}\mu_h\right\}\phi(\boldsymbol{t})\bigg|_{t=0}$$
$$=\mu_h.$$

二阶矩是

$$\mathrm{E}(X_hX_j)=\frac{1}{\mathrm{i}^2}\frac{\partial^2\phi}{\partial t_h\partial t_j}\bigg|_{t=0} \tag{22}$$
$$=\frac{1}{\mathrm{i}^2}\left\{\left(-\sum_k \sigma_{hk}t_k+\mathrm{i}\mu_h\right)\left(-\sum_j \sigma_{kj}t_k+\mathrm{i}\mu_j\right)-\sigma_{hj}\right\}\phi(\boldsymbol{t})\bigg|_{t=0}$$
$$=\sigma_{hj}+\mu_h\mu_j.$$

因此

$$\mathrm{Var}(X_i)=\mathrm{E}[(X_i-\mu_i)^2]=\sigma_{ii}, \tag{23}$$

$$\mathrm{Cov}(X_i,X_j)=\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)]=\sigma_{ij}. \tag{24}$$

任何关于均值的三阶矩是

$$\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)]=0. \tag{25}$$

关于均值的四阶矩是

$$\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)]=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}. \tag{26}$$

每个奇数阶矩都是 0.

定义 2.6.3 *如果一个分布的所有的矩都存在, 则其半不变量是*

$$\ln\,\phi(\boldsymbol{t})=\sum_{s_1,\cdots,s_p=0}^{\infty}\kappa_{s_1\cdots s_p}\frac{(\mathrm{i}t_1)^{s_1}\cdots(\mathrm{i}t_p)^{s_p}}{s_1!\cdots s_p!} \tag{27}$$

中的系数 κ.

在多元正态分布情形下, $\kappa_{10\cdots0}=\mu_1,\cdots,\kappa_{0\cdots01}=\mu_p,\kappa_{20\cdots0}=\sigma_{11},\cdots,\kappa_{0\cdots02}=\sigma_{pp},\kappa_{110\cdots0}=\sigma_{12},\cdots$. 当 $\sum s_i>2$ 时, 其半不变量为 0.

2.7 椭球等高分布

2.7.1 球形和椭球等高分布

我们在 2.3 节末指出, 均值为 $\boldsymbol{\mu}$ 且协方差阵为 $\boldsymbol{\Sigma}$ 的多元正态分布在同心椭球

$$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=k. \tag{1}$$

上的密度是常数.

具有这个性质的一族分布称为*椭球等高分布族*, 其密度为

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g\left[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})\right], \tag{2}$$

其中 $\boldsymbol{\Lambda}$ 是一个正矩定阵, $g(\cdot)\geqslant 0$, 且

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(\boldsymbol{y}'\boldsymbol{y})\mathrm{d}y_1\cdots\mathrm{d}y_p=1. \tag{3}$$

如果 $\boldsymbol{C}$ 是非奇异矩阵且 $\boldsymbol{C}'\boldsymbol{\Lambda}^{-1}\boldsymbol{C}=\boldsymbol{I}$, 变换 $\boldsymbol{x}-\boldsymbol{\nu}=\boldsymbol{C}\boldsymbol{y}$ 将密度 (2) 变为密度 $g(\boldsymbol{y}'\boldsymbol{y})$. $g(\boldsymbol{y}'\boldsymbol{y})$ 的常密度等高线是中心在原点的球面. 这样的一族分布称为**球面等高分布**. 椭球等高分布不必有密度, 但是只有有密度的分布才有统计意义.

球面等高密度可以用极坐标表示, 做变换

$$\begin{aligned}
y_1&=r\sin\theta_1, \\
y_2&=r\cos\theta_1\sin\theta_2,\\
y_3&=r\cos\theta_1\cos\theta_2\sin\theta_3,\\
&\vdots\\
y_{p-1}&=r\cos\theta_1\cos\theta_2\cdots\cos\theta_{p-2}\sin\theta_{p-2},\\
y_p&=r\cos\theta_1\cos\theta_2\cdots\cos\theta_{p-2}\cos\theta_{p-1},
\end{aligned} \tag{4}$$

其中 $-\frac{1}{2}\pi<\theta_i\leqslant\frac{1}{2}\pi, i=1,\cdots,p-2, -\pi<\theta_{p-1}\leqslant\pi$, 且 $0\leqslant r<\infty$. 注意 $\boldsymbol{y}'\boldsymbol{y}=r^2$. 变换 (4) 的雅可比行列式是 $r^{p-1}\cos^{p-2}\theta_1\cos^{p-3}\theta_2\cdots\cos\theta_{p-2}$. 见习题 7.1. 如果 $g(\boldsymbol{y}'\boldsymbol{y})$ 是 $\boldsymbol{Y}$ 的密度, 则 $R,\Theta_1,\cdots,\Theta_{p-1}$ 的密度是

$$r^{p-1}\cos^{p-2}\theta_1\cos^{p-3}\theta_2\cdots\cos\theta_{p-2}g(r^2). \tag{5}$$

注意 $R,\Theta_1,\cdots,\Theta_{p-1}$ 是独立分布的. 因为

$$\int_{-\pi/2}^{\pi/2}\cos^{h-1}\theta\mathrm{d}\theta=\frac{\Gamma(\frac{1}{2}h)\Gamma(\frac{1}{2})}{\Gamma[\frac{1}{2}(h+1)]} \tag{6}$$

(习题 7.2), 所以 R 的边缘密度是

$$C(p)g(r^2)r^{p-1}, \tag{7}$$

其中

$$\begin{aligned}
C(p)&=\frac{2\pi^{\frac{1}{2}p}}{\Gamma(\frac{1}{2}p)}\\
&=\int_{-\pi}^{\pi}\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2}\cos^{p-2}\theta_1\cos^{p-3}\theta_2\cdots\cos\theta_{p-2}\mathrm{d}\theta_1\cdots\mathrm{d}\theta_{p-2}\mathrm{d}\theta_{p-1}.
\end{aligned} \tag{8}$$

Θ_i 的边缘密度是 $\Gamma[\frac{1}{2}(p-i)]\cos^{p-i-1}\theta/\{\Gamma(\frac{1}{2})\Gamma[\frac{1}{2}(p-i-1)]\}, i=1,\cdots,p-2$, Θ_{p-1} 的边缘密度是 $1/(2\pi)$.

在正态情形 $N(\boldsymbol{0},\boldsymbol{I})$ 下, $\boldsymbol{Y}$ 的密度是

$$g(\boldsymbol{y}'\boldsymbol{y})=(2\pi)^{-\frac{1}{2}p}\exp\left(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}\right),$$

且 $R=(\boldsymbol{Y}'\boldsymbol{Y})^{\frac{1}{2}}$ 的密度是 $r^{p-1}\exp(-\frac{1}{2}r^2)/[2^{\frac{1}{2}p-1}\Gamma(\frac{1}{2}p)]$. $r^2=\nu$ 的密度是 $\nu^{\frac{1}{2}p-1}\mathrm{e}^{-\frac{1}{2}\nu}/[2^{\frac{1}{2}p}\Gamma(\frac{1}{2}p)]$. 这是自由度为 p 的 χ^2 密度.

常数 $C(p)$ 是 p 维空间中单位半径球面的曲面面积. 称坐标为 $\sin\Theta_1$, $\cos\Theta_1\sin\Theta_2,\cdots$, $\cos\Theta_1\cos\Theta_2\cdots\cos\Theta_{p-1}$ 的随机向量 $\boldsymbol{U}$ 为*在单位球面上均匀分布的*, 若 $\Theta_1,\cdots,\Theta_{p-1}$ 相互独立, 且除 Θ_{p-1} 在 $(-\pi,\pi)$ 上均匀分布, 其他变量都在 $(-\pi/2,\pi/2)$ 上均匀分布. (这是一个最简单的没有密度的球面等高分布的例子.) 密度为 $g(\boldsymbol{y}'\boldsymbol{y})$ 的 $\boldsymbol{Y}$ 的一个随机表示是

$$\boldsymbol{Y}\overset{d}{=}R\boldsymbol{U}, \tag{9}$$

其中 R 的密度是 (7).

因为 $\Theta_1,\cdots,\Theta_{p-1}$ 中每一个的密度都是偶函数, 所以

$$\mathrm{E}(\boldsymbol{U})=\boldsymbol{0}. \tag{10}$$

因为 R 和 $\boldsymbol{U}$ 独立, 所以如果 $\mathrm{E}(R)<\infty$, 则有

$$\mathrm{E}(\boldsymbol{Y})=\boldsymbol{0}. \tag{11}$$

进一步, 如果 $\mathrm{E}(R^2)<\infty$, 则有

$$\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')=\mathrm{E}(R^2)\mathrm{E}(\boldsymbol{U}\boldsymbol{U}'). \tag{12}$$

因为 $\sum_{i=1}^p U_i^2=1$, 由对称性知 $\mathrm{E}(U_1^2)=\cdots=\mathrm{E}(U_p^2)=1/p$. 又由对称性知 $\mathrm{E}(U_1U_2)=\mathrm{E}(U_1U_3)=\cdots=\mathrm{E}(U_{p-1}U_p)$. 特别地, $\mathrm{E}(U_1U_2)=\mathrm{E}(\sin\Theta_1\cos\Theta_1\sin\Theta_2)$, 被积函数是 θ_1 和 θ_2 的奇函数. 因此, $\mathrm{E}(U_iU_j)=0, i\neq j$. 总之,

$$\mathrm{E}(\boldsymbol{U}\boldsymbol{U}')=(1/p)\boldsymbol{I}_p, \tag{13}$$

$$\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')=(1/p)\mathrm{E}(R^2)\boldsymbol{I}_p \tag{14}$$

(如果 $\mathrm{E}(R^2)<\infty$).

球面等高分布族的显著特征是, 对任意正交矩阵 $\boldsymbol{O}$, 有 $\boldsymbol{O}\boldsymbol{Y}\overset{d}{=}\boldsymbol{Y}$.

定理 2.7.1 *如果 $\boldsymbol{Y}$ 有密度 $g(\boldsymbol{y}'\boldsymbol{y})$, 则 $\boldsymbol{Z}=\boldsymbol{O}\boldsymbol{Y}$ 有密度 $g(\boldsymbol{z}'\boldsymbol{z})$, 其中 $\boldsymbol{O}'\boldsymbol{O}=\boldsymbol{I}$.*

证明 变换 $\boldsymbol{z}=\boldsymbol{O}\boldsymbol{y}$ 的雅可比行列式是 1. ■

我们可将 $\boldsymbol{Y}$ 为球面等高分布的定义推广到任何具有性质 $\boldsymbol{O}\boldsymbol{Y}\overset{d}{=}\boldsymbol{Y}$ 的分布.

推论 2.7.1 *如果 $\boldsymbol{Y}$ 是球面等高分布的, 且其具有随机表示 $\boldsymbol{Y}\overset{d}{=}R\boldsymbol{U}$, 其中 $R^2=\boldsymbol{Y}'\boldsymbol{Y}$, 则 $\boldsymbol{U}$ 也是球面等高分布的.*

证明 如果 $\boldsymbol{Z}=\boldsymbol{O}\boldsymbol{Y}$, 则 $\boldsymbol{Z}\overset{d}{=}\boldsymbol{Y}$ 且 $\boldsymbol{Z}$ 有随机表示 $\boldsymbol{Z}=S\boldsymbol{V}$, 其中 $S^2=\boldsymbol{Z}'\boldsymbol{Z}$, 于是 $S=R$ 且 $\boldsymbol{V}=\boldsymbol{O}\boldsymbol{U}\overset{d}{=}\boldsymbol{U}$. ■

$\boldsymbol{X}=\boldsymbol{\nu}+\boldsymbol{C}\boldsymbol{Y}$ 的密度是 (2). 从 (11) 和 (14) 中, 我们可以得到下面的定理.

定理 2.7.2 *如果 $\boldsymbol{X}$ 有密度 (2) 且 $\mathrm{E}(\boldsymbol{R}^2)<\infty$, 那么*

$$\mathrm{E}(\boldsymbol{X})=\boldsymbol{\mu}=\boldsymbol{\nu},\quad \mathrm{Cov}(\boldsymbol{X})=\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})']=\boldsymbol{\Sigma}=(1/p)\mathrm{E}(\boldsymbol{R}^2\boldsymbol{\Lambda}). \tag{15}$$

实际上, 如果 $\mathrm{E}(\boldsymbol{R}^m)<\infty$, 则 $\boldsymbol{X}$ 的 $h(h\leqslant m)$ 阶矩是 $\mathrm{E}[(X_1-\mu_1)^{h_1}\cdots(X_p-\mu_p)^{h_p}]=\mathrm{E}(Z_1^{h_1}\cdots Z_p^{h_p})\mathrm{E}(\boldsymbol{R}^h)/\mathrm{E}(\chi_p^2)^{\frac{1}{2}h}$, 其中 $\boldsymbol{Z}$ 有分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$ 且 $h=h_1+\cdots+h_p$.

定理 2.7.3 如果 $\boldsymbol{X}$ 有密度 (2), $\mathrm{E}(R^2)<\infty$ 且对所有的 $c>0$ 有 $f[c\mathrm{Cov}(\boldsymbol{X})]=f[\mathrm{Cov}(\boldsymbol{X})]$, 则 $f[\mathrm{Cov}(\boldsymbol{X})]=f(\boldsymbol{\Sigma})$.

特别地, $\rho_{ij}(\boldsymbol{X})=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}=\lambda_{ij}/\sqrt{\lambda_{ii}\lambda_{jj}}$, 其中 $\boldsymbol{\Sigma}=(\sigma_{ij})$ 且 $\boldsymbol{\Lambda}=(\lambda_{ij})$.

2.7.2 线性组合的分布, 边缘分布

首先, 我们考虑密度为 $g(\boldsymbol{y}'\boldsymbol{y})$ 的球面等高分布. 设 $\boldsymbol{y}'=(\boldsymbol{y}_1',\boldsymbol{y}_2')$, 其中 $\boldsymbol{y}_1$ 和 $\boldsymbol{y}_2$ 分别有 q 个和 $p-q$ 个分量. $\boldsymbol{y}_2$ 的边缘分布是

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(\boldsymbol{y}_1'\boldsymbol{y}_1+\boldsymbol{y}_2'\boldsymbol{y}_2)\mathrm{d}y_1\cdots\mathrm{d}y_q. \tag{16}$$

将 $\boldsymbol{y}_1$ 用极坐标 (4) 表示, 其中用 r_1 代替 r, 用 q 代替 p. 则 $\boldsymbol{y}_2$ 的边缘密度是

$$g_2(\boldsymbol{y}_2'\boldsymbol{y}_2)=C(q)\int_0^{\infty}g(r_1^2+\boldsymbol{y}_2'\boldsymbol{y}_2)r_1^{q-1}\mathrm{d}r_1. \tag{17}$$

这个表达式说明了 $\boldsymbol{y}_2$ 的边缘分布也有球面等高分布密度.

现在考虑一个密度为 (2) 的向量 $\boldsymbol{X}'=(\boldsymbol{X}^{(1)'},\boldsymbol{X}^{(2)'})$. 如果 $\mathrm{E}(R^2)<\infty$, 则 $\boldsymbol{X}$ 的协方差阵 (15) 可被划分为 2.4 节的 (14) 那样. 令 $\boldsymbol{Z}^{(1)}=\boldsymbol{X}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}^{(2)}=\boldsymbol{X}^{(1)}-\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{X}^{(2)}$, $\boldsymbol{Z}^{(2)}=\boldsymbol{X}^{(2)}$, $\boldsymbol{\tau}^{(1)}=\boldsymbol{\nu}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\nu}^{(2)}=\boldsymbol{\nu}^{(1)}-\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\nu}^{(2)}$, $\boldsymbol{\tau}^{(2)}=\boldsymbol{\nu}^{(2)}$. 则 $\boldsymbol{Z}'=(\boldsymbol{Z}^{(1)'},\boldsymbol{Z}^{(2)'})$ 的密度是

$$\begin{aligned}&|\boldsymbol{\Lambda}_{11\cdot2}|^{-\frac{1}{2}}|\boldsymbol{\Lambda}_{22}|^{-\frac{1}{2}}g[(\boldsymbol{z}^{(1)}-\boldsymbol{\tau}^{(1)})'\boldsymbol{\Lambda}_{11\cdot2}(\boldsymbol{z}^{(1)}-\boldsymbol{\tau}^{(1)})\\&\qquad+(\boldsymbol{z}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}'(\boldsymbol{z}^{(2)}-\boldsymbol{\nu}^{(2)})].\end{aligned} \tag{18}$$

注意到, 尽管 $\boldsymbol{Z}^{(1)}$ 和 $\boldsymbol{Z}^{(2)}$ 可能不独立, 但它们一定不相关. 设 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 分别是满足 $\boldsymbol{C}_1\boldsymbol{\Lambda}_{11\cdot2}^{-1}\boldsymbol{C}_1'=\boldsymbol{I}_q$ 和 $\boldsymbol{C}_2\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{C}_2'=\boldsymbol{I}_{p-q}$ 的 $q\times q$ 矩阵和 $(p-q)\times(p-q)$ 矩阵. 定义 $\boldsymbol{y}^{(1)}$ 和 $\boldsymbol{y}^{(2)}$ 分别为 $\boldsymbol{z}^{(1)}-\boldsymbol{\tau}^{(1)}=\boldsymbol{C}_1\boldsymbol{y}^{(1)}$ 和 $\boldsymbol{z}^{(2)}-\boldsymbol{\nu}^{(2)}=\boldsymbol{C}_2\boldsymbol{y}^{(2)}$. 则 $\boldsymbol{Y}^{(1)}$ 和 $\boldsymbol{Y}^{(2)}$ 有密度 $g(\boldsymbol{y}^{(1)'}\boldsymbol{y}^{(1)}+\boldsymbol{y}^{(2)'}\boldsymbol{y}^{(2)})$. $\boldsymbol{Y}^{(2)}$ 的边缘密度是 (17), 并且 $\boldsymbol{X}^{(2)}=\boldsymbol{Z}^{(2)}$ 的边缘密度是

$$\begin{aligned}&|\boldsymbol{\Lambda}_{22}|^{-\frac{1}{2}}g_2[(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]\\&=C(q)\int_0^{\infty}g[r_1^2+(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]r_1^{q-1}\mathrm{d}r_1.\end{aligned} \tag{19}$$

$\boldsymbol{Y}_2$ 的矩可由 $\boldsymbol{Y}$ 的矩计算得到.

定理 2.4.1 在椭球等高分布上的推广如下: 设 p 维向量 $\boldsymbol{X}$ 有密度 (2). 则对于非奇异矩阵 $\boldsymbol{C}$, $\boldsymbol{Y}=\boldsymbol{C}\boldsymbol{X}$ 有密度 $|\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}'|^{-\frac{1}{2}}g_2[(\boldsymbol{x}-\boldsymbol{C}\boldsymbol{\nu})'(\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}')^{-1}(\boldsymbol{x}-\boldsymbol{C}\boldsymbol{\nu})]$.

定理 2.4.4 在椭球等高分布上的推广如下: 如果 $\boldsymbol{X}$ 有密度 (2), 则 $\boldsymbol{Z}=\boldsymbol{D}\boldsymbol{X}$ 有密度

$$|\boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}'|^{-\frac{1}{2}}g_2\left[(\boldsymbol{z}-\boldsymbol{D}\boldsymbol{\nu})'(\boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}')^{-1}(\boldsymbol{z}-\boldsymbol{D}\boldsymbol{\nu})\right], \tag{20}$$

其中 $\boldsymbol{D}$ 是一个秩为 $q\leqslant p$ 的 $q\times p$ 矩阵, 且 g_2 由 (17) 给出.

我们也可用表示 (9) 的方法来刻画边缘分布. 考虑

$$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}^{(1)}\\ \boldsymbol{Y}^{(2)}\end{pmatrix}\overset{d}{=}R\boldsymbol{U}=R\begin{pmatrix}\boldsymbol{U}^{(1)}\\ \boldsymbol{U}^{(2)}\end{pmatrix},\tag{21}$$

其中 $\boldsymbol{Y}^{(1)}$ 和 $\boldsymbol{U}^{(1)}$ 有 q 个分量, $\boldsymbol{Y}^{(2)}$ 和 $\boldsymbol{U}^{(2)}$ 有 $p-q$ 个分量. 则 $R_2^2=\boldsymbol{Y}^{(2)\prime}\boldsymbol{Y}^{(2)}$ 和 $R^2\boldsymbol{U}^{(2)\prime}\boldsymbol{U}^{(2)}$ 同分布, 且

$$\boldsymbol{U}^{(2)\prime}\boldsymbol{U}^{(2)}=\frac{\boldsymbol{U}^{(2)\prime}\boldsymbol{U}^{(2)}}{\boldsymbol{U}'\boldsymbol{U}}\overset{d}{=}\frac{\boldsymbol{Y}^{(2)\prime}\boldsymbol{Y}^{(2)}}{\boldsymbol{Y}'\boldsymbol{Y}}.\tag{22}$$

当 $\boldsymbol{Y}\sim N(\boldsymbol{0},\boldsymbol{I}_p)$ 时, (22) 服从贝塔分布, 即 $B(p-q,q)$, 其密度为

$$\frac{\Gamma(p/2)}{\Gamma(q/2)\Gamma[(p-q)/2]}z^{\frac{1}{2}(p-q)-1}(1-z)^{\frac{1}{2}q-1},\qquad 0\leqslant z\leqslant 1.\tag{23}$$

继而总之,

$$\boldsymbol{Y}^{(2)}\overset{d}{=}R_2\boldsymbol{V},\tag{24}$$

其中, $R_2^2\overset{d}{=}R^2b, b\sim B(p-q,q)$, $\boldsymbol{V}$ 是 p_2 维空间里 $\boldsymbol{v}'\boldsymbol{v}=1$ 的均匀分布, 并且 R^2, b 和 $\boldsymbol{V}$ 是相互独立的. 所有边缘分布都是椭球等高的.

2.7.3 条件分布和多重相关系数

当 $\boldsymbol{y}=(\boldsymbol{y}_1',\boldsymbol{y}_2')'$ 有球面密度 $g(\boldsymbol{y}'\boldsymbol{y})$ 时, 给定 $\boldsymbol{y}_2$ 时 $\boldsymbol{y}_1$ 的条件分布的密度是

$$\frac{g(\boldsymbol{y}_1'\boldsymbol{y}_1+\boldsymbol{y}_2'\boldsymbol{y}_2)}{g_2(\boldsymbol{y}_2'\boldsymbol{y}_2)}=\frac{g(\boldsymbol{y}_1'\boldsymbol{y}_1+r_2^2)}{g_2(r_2^2)},\tag{25}$$

其中, 边缘密度 $g_2(\boldsymbol{y}_2'\boldsymbol{y}_2)$ 由 (17) 给出, $r_2^2=\boldsymbol{y}_2'\boldsymbol{y}_2$. 从 $\boldsymbol{y}_1$ 的角度来看, (25) 是一个球面等高分布 (依赖于 r_2^2).

现在考虑 $\boldsymbol{X}=(\boldsymbol{X}_1',\boldsymbol{X}_2')'$, 其密度为 (2). 给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件密度是

$$\begin{aligned}&|\boldsymbol{\Lambda}_{11\cdot 2}|^{-\frac{1}{2}}g\{[(\boldsymbol{x}^{(1)}-\boldsymbol{\nu}^{(1)})'-(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{B}']\boldsymbol{\Lambda}_{11\cdot 2}^{-1}[\boldsymbol{x}^{(1)}-\boldsymbol{\nu}^{(1)}-\boldsymbol{B}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]\\&+(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})\}\div g_2[(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]\\=&|\boldsymbol{\Lambda}_{11\cdot 2}|^{-\frac{1}{2}}g\{[\boldsymbol{x}^{(1)}-\boldsymbol{\nu}^{(1)}-\boldsymbol{B}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]'\boldsymbol{\Lambda}_{11\cdot 2}^{-1}[\boldsymbol{x}^{(1)}-\boldsymbol{\nu}^{(1)}-\boldsymbol{B}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})]+r_2^2\}\\&\div g_2(r_2^2),\end{aligned}\tag{26}$$

其中 $r_2^2=(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})'\boldsymbol{\Lambda}_{22}^{-1}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})$ 且 $\boldsymbol{B}=\boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}$. 视 $\boldsymbol{x}^{(1)}-\boldsymbol{\nu}^{(1)}-\boldsymbol{B}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)})$ 是 $\boldsymbol{x}^{(1)}$ 的一个函数, 密度 (26) 是椭球等高的. 如果 (25) 中 $\mathrm{E}(R_1^2|\boldsymbol{y}_2'\boldsymbol{y}_2=r_2^2)<\infty$, 其中 $R_1^2=\boldsymbol{Y}_1'\boldsymbol{Y}_1$, 则给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件均值是

$$\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)})=\boldsymbol{\nu}^{(1)}+\boldsymbol{B}(\boldsymbol{x}^{(2)}-\boldsymbol{\nu}^{(2)}).\tag{27}$$

同样, 条件协方差阵是 $(\mathrm{E}(r_2^2/q))\boldsymbol{\Lambda}_{11\cdot 2}$. 由此可知, 当 $(\sigma_{ij\cdot q+1,\cdots,p})=\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 且 $\boldsymbol{\Sigma}$ 是如上给定的参数矩阵时, 偏相关系数的定义 2.5.2 成立.

当 $\mathrm{E}(R^2)<\infty$ 时, 定理 2.5.2、定理 2.5.3 和定理 2.5.4 对任意椭球等高分布都成立.

2.7.4 特征函数, 矩

一个球面等高分布随机向量 $\boldsymbol{Y}$ 的特征函数 $\mathrm{E}e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{Y}}$ 具有在正交变换下不变的性质, 即

$$\begin{aligned}\mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{O}\boldsymbol{Y}}) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{O}\boldsymbol{y}} g(\boldsymbol{y}'\boldsymbol{y})\mathrm{d}y_1\cdots\mathrm{d}y_p \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{z}} g(\boldsymbol{z}'\boldsymbol{z})\mathrm{d}z_1\cdots\mathrm{d}z_p \\ &= \mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}}),\end{aligned} \tag{28}$$

其中 $\boldsymbol{Z}=\boldsymbol{O}\boldsymbol{Y}$ 也有密度 $g(\boldsymbol{y}'\boldsymbol{y})$. 对任意正交矩阵 $\boldsymbol{O}$, 等式 (28) 说明了 $\mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{Z}})$ 是 $\boldsymbol{t}'\boldsymbol{t}$ 的函数. 记

$$\mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{Y}}) = \phi(\boldsymbol{t}'\boldsymbol{t}). \tag{29}$$

则对 $\boldsymbol{X}=\boldsymbol{\mu}+\boldsymbol{C}\boldsymbol{Y}$,

$$\begin{aligned}\mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{X}}) &= e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\mathrm{E}(e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{C}\boldsymbol{Y}}) \\ &= e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\phi(\boldsymbol{t}'\boldsymbol{C}\boldsymbol{C}'\boldsymbol{t}) \\ &= e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\phi(\boldsymbol{t}'\boldsymbol{\Lambda}\boldsymbol{t}),\end{aligned} \tag{30}$$

其中 $\boldsymbol{\Lambda}=\boldsymbol{C}\boldsymbol{C}'$. 相反地, 任何特征函数为 $e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\phi(\boldsymbol{t}'\boldsymbol{\Lambda}\boldsymbol{t})$ 形式的随机向量 $\boldsymbol{X}$, 其密度为 (2).

椭球等高分布的随机向量 $\boldsymbol{X}$ 的矩可以从其特征函数 $e^{\mathrm{i}\boldsymbol{t}'\boldsymbol{\mu}}\phi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})$ 得到, 或从表示 $\boldsymbol{X}=\boldsymbol{\mu}+R\boldsymbol{C}\boldsymbol{U}$ 得到, 其中 $\boldsymbol{C}'\boldsymbol{\Lambda}^{-1}\boldsymbol{C}=\boldsymbol{I}$. 注意

$$\mathrm{E}(R^2) = C(p)\int_0^{\infty} r^{p+1}g(r^2)\mathrm{d}r = -2p\phi'(0), \tag{31}$$

$$\mathrm{E}(R^4) = C(p)\int_0^{\infty} r^{p+3}g(r^2)\mathrm{d}r = 4p(p+2)\phi''(0). \tag{32}$$

考虑 $\boldsymbol{Y}=R\boldsymbol{U}$ 的高阶矩. R 的奇数阶矩是 0, 因此 $\boldsymbol{Y}$ 的奇数阶矩也是 0.

我们有

$$\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)] = 0. \tag{33}$$

实际上, 所有 $\boldsymbol{X}-\boldsymbol{\mu}$ 的奇数阶矩都是 0.

考虑 $\mathrm{E}(U_iU_jU_kU_l)$. 因为 $\boldsymbol{U}'\boldsymbol{U}=1$, 所以

$$1=\sum_{i,j=1}^{p}\mathrm{E}(U_i^2U_j^2) = p\mathrm{E}(U_1^4)+p(p-1)\mathrm{E}(U_1^2U_2^2). \tag{34}$$

由 $\mathrm{E}(\sin^4\Theta_1)$ 的积分得 $\mathrm{E}(U_1^4)=3/[p(p+2)]$, 则 (34) 说明 $\mathrm{E}(U_1^2U_2^2)=1/[p(p+2)]$. 因此 $\mathrm{E}(Y_i^4)=3\mathrm{E}(R^4)/[p(p+2)]$ 且 $\mathrm{E}(Y_1^2Y_2^2)=\mathrm{E}(R^4)/[p(p+2)]$. 除 $i=j=k=l$ 或 $i=j\neq k=l$ 或 $i=k\neq j=l$ 或 $i=l\neq j=k$ 之外, 我们有 $\mathrm{E}(U_iU_jU_kU_l)=0$. 综上, $\mathrm{E}(U_iU_jU_kU_l)=(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk})/[p(p+2)]$. $\boldsymbol{X}$ 的四阶矩是

$$\begin{aligned}&\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)] \\ =&\frac{\mathrm{E}(R^4)}{p(p+2)}(\lambda_{ij}\lambda_{kl}+\lambda_{ik}\lambda_{jl}+\lambda_{il}\lambda_{jk}) \\ =&\frac{\mathrm{E}(R^4)}{(\mathrm{E}(R^2))^2}\frac{p}{p+2}(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}).\end{aligned} \tag{35}$$

经标准差标准化后的随机向量 $\boldsymbol{X}$ 的第 i 个分量的四阶半不变量是

$$\begin{aligned}\frac{\mathrm{E}(X_i-\mu_i)^4}{[\mathrm{E}(X_i-\mu_i)^2]^2}-3=&\frac{\frac{3\mathrm{E}(R^4)}{p(p+2)}-3\left(\frac{\mathrm{E}(R^2)}{p}\right)^2}{\left(\frac{\mathrm{E}(R^2)}{p}\right)^2} \\ =&3\left[\frac{\mathrm{E}(R^4)}{(\mathrm{E}(R^2))^2}\frac{p}{p+2}-1\right]=3\left[\frac{\phi''(0)}{[\phi'(0)]^2}-1\right] \\ =&3\kappa.\end{aligned} \tag{36}$$

这就是所谓的**峰度**. (注意 κ 是 $\frac{1}{3}\mathrm{E}\{(X_i-\mu_i)^4/[\mathrm{E}(X_i-\mu_i)^2]^2\}-1$.) $\boldsymbol{X}$ 的每个分量的标准四阶半不变量都是 3κ. X_i,X_j,X_k 和 X_l 的四阶半不变量是

$$\begin{aligned}\kappa_{ijkl}=&\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)]-(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}) \\ =&\kappa(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}).\end{aligned} \tag{37}$$

对于正态分布 $\kappa=0$. 四阶矩可以记为

$$\begin{aligned}&\mathrm{E}[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)] \\ =&(1+\kappa)(\sigma_{ij}\delta_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}).\end{aligned} \tag{38}$$

更多关于椭球等高分布的细节可以参考 Fang and Zhang(1990).

椭球等高分布族推广了正态分布, 它更灵活. 峰度不一定为 0. 密度为 $|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})]$ 的典型"钟形曲面"相比正态分布可以有或多或少的尖峰. 在下面的小节里将给出一些例子.

2.7.5 例子

(1) *多元 t 分布*. 假设 $\boldsymbol{Z}\sim N(\boldsymbol{0},\boldsymbol{I}_p)$, $ms^2\stackrel{d}{=}\chi^2_m$, 且 $\boldsymbol{Z}$ 和 s^2 独立. 定义 $\boldsymbol{Y}=(1/s)\boldsymbol{Z}$. 则 $\boldsymbol{Y}$ 的密度是

$$\frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})m^{p/2}\pi^{p/2}}\left(1+\frac{\boldsymbol{y}'\boldsymbol{y}}{m}\right)^{-\frac{m+p}{2}}, \tag{39}$$

且

$$\frac{R^2}{p}=\frac{||\boldsymbol{Y}||^2}{p}\sim F_{p,m}=\frac{m}{p}\frac{\chi^2_p}{\chi^2_m}. \tag{40}$$

如果 $\boldsymbol{X}=\boldsymbol{\mu}+\boldsymbol{C}\boldsymbol{Y}$, 则 $\boldsymbol{X}$ 的密度是

$$\frac{\Gamma(\frac{m+p}{2})}{\Gamma(\frac{m}{2})m^{p/2}\pi^{p/2}}|\boldsymbol{\Lambda}|^{-\frac{1}{2}}\left[1+\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{m}\right]^{-\frac{1}{2}(m+p)}. \tag{41}$$

(2) **污染的正态分布**. 污染的正态分布是两个协方差阵等比、均值向量相同的正态分布的组合. 其密度可以记为

$$(1-\varepsilon)\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Lambda}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))} \\ +\varepsilon\frac{1}{(2\pi)^{p/2}|c\boldsymbol{\Lambda}|^{\frac{1}{2}}}\mathrm{e}^{-(1/2c)(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}, \tag{42}$$

其中 $c>0,\ 0\leqslant\varepsilon\leqslant1$. 通常 ε 比较小, c 比较大.

(3) **混合正态分布**. 设 $w(\nu)$ 是 $0\leqslant\nu\leqslant\infty$ 上的一个累积分布函数. 混合正态密度的定义为

$$\int_0^\infty n\left(\boldsymbol{x}|\boldsymbol{\mu},\frac{1}{\nu^2}\boldsymbol{\Sigma}\right)\mathrm{d}w(\nu), \tag{43}$$

它是一个椭球等高密度. 若随机向量 $\boldsymbol{X}$ 有如上密度, 则它可以表示为 $\boldsymbol{X}=w\boldsymbol{Z}$, 其中 $\boldsymbol{Z}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 和 $w\sim w(w)$ 独立.

Fang, Kotz and Ng (1990) 讨论了 (43) 并给出了一些其他椭球等高分布的例子.

习　　题

2.1　(2.2 节)　设 $f(x,y)=\begin{cases}1, & 0\leqslant x\leqslant1, 0\leqslant y\leqslant1,\\ 0, & \text{其他}.\end{cases}$

求 (a) $F(x,y)$; (b) $F(x)$; (c) $f(x)$; (d) $f(x|y)$; [注意, 如果 $f(x_0,y_0)=0$, 则 $f(x_0|y_0)=0$.] (e) $\mathrm{E}(X^nY^m)$; (f) 证明 X 和 Y 独立.

2.2　(2.2 节)　设 $f(x,y)=\begin{cases}2, & 0\leqslant y\leqslant x\leqslant1,\\ 0, & \text{其他}\end{cases}$

求 (a) $F(x,y)$; (b) $F(x)$; (c) $f(x)$; (d) $G(y)$; (e) $g(y)$; (f) $f(x|y)$; (g) $f(y|x)$; (h) $\mathrm{E}(X^nY^m)$; (i) X 和 Y 独立么?

2.3　(2.2 节)　设当 $x^2+y^2\leqslant k^2$ 时 $f(x,y)=C$, 否则为 0. 证明 $C=1/(\pi k^2)$, $\mathrm{E}(X)=\mathrm{E}(Y)=0$, $\mathrm{E}(X^2)=\mathrm{E}(Y^2)=k^2/4$ 且 $\mathrm{E}(XY)=0$. 则 X 和 Y 独立么?

2.4　(2.2 节)　设 $F(x_1,x_2)$ 是 X_1,X_2 的联合 cdf , 且设 $F_i(x_i)$ 是 X_i 的边缘 cdf , $i=1,2$. 证明: 如果 $F_i(x_i)$ 是连续的, $i=1,2$, 则 $F(x_1,x_2)$ 也是连续的.

2.5　(2.2 节)　证明: 如果集合 $X_1,\cdots,X_r$ 与集合 $X_{r+1},\cdots,X_p$ 独立, 则

$$\mathrm{E}[g(X_1,\cdots,X_r)h(X_{r+1},\cdots,X_p)]=\mathrm{E}[g(X_1,\cdots,X_r)]\mathrm{E}[h(X_{r+1},\cdots,X_p)].$$

2.6　(2.3 节)　画椭圆 $f(x,y)=0.06$, 其中 $f(x,y)$ 是一个二元正态密度且

(a) $\mu_x=1,\ \mu_y=2,\ \sigma_x^2=1,\ \sigma_y^2=1,\ \rho_{xy}=0$,

(b) $\mu_x=0,\ \mu_y=0,\ \sigma_x^2=1,\ \sigma_y^2=1,\ \rho_{xy}=0$,

(c) $\mu_x=0,\ \mu_y=0,\ \sigma_x^2=1,\ \sigma_y^2=1,\ \rho_{xy}=0.2$,

(d) $\mu_x=0,\ \mu_y=0,\ \sigma_x^2=1,\ \sigma_y^2=1,\ \rho_{xy}=0.8$,

(e) $\mu_x=0$, $\mu_y=0$, $\sigma_x^2=4$, $\sigma_y^2=1$, $\rho_{xy}=0.8$.

2.7 (2.3 节) 求 $\boldsymbol{b}$ 和 $\boldsymbol{A}$, 使得下面的密度可以写成 (23) 的形式. 并求 u_x, u_y, σ_x, σ_y 和 ρ_{xy}.

(a) $\frac{1}{2\pi}\exp\{-\frac{1}{2}[(x-1)^2+(y-2)^2]\}$.

(b) $\frac{1}{2.4\pi}\exp\left(-\frac{x^2/4-1.6xy/2+y^2}{0.72}\right)$.

(c) $\frac{1}{2\pi}\exp[-\frac{1}{2}(x^2+y^2+4x-6y+13)]$.

(d) $\frac{1}{2\pi}\exp[-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)]$.

2.8 (2.3 节) 对于习题 **2.7** 的每一个矩阵 $\boldsymbol{A}$, 求 $\boldsymbol{C}$, 使得 $\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}=\boldsymbol{I}$.

2.9 (2.3 节) 设 $\boldsymbol{b}=\boldsymbol{0}$,

$$\boldsymbol{A}=\begin{pmatrix}7&3&2\\3&4&1\\2&1&2\end{pmatrix}.$$

(a)写出密度 (23). (b) 求 $\boldsymbol{\Sigma}$.

2.10 (2.3 节) 证明 2.3 节中, 相应于变换 $y_1=(z_1+z_2)/\sqrt{2}, y_2=(z_1-z_2)/\sqrt{2}$, (55) 的主轴分别沿 45° 和 135° 方向, 长度分别为 $2\sqrt{c(1+\rho)}$ 和 $2\sqrt{c(1-\rho)}$.

2.11 (2.3 节) 假设标量随机变量 $X_1,\cdots,X_n$ 是相互独立的且其密度只依赖于 $x_1^2+\cdots+x_n^2$. 证明 X_i 服从均值为 0、方差相等的正态分布. 在证明中指出对密度要求的最弱条件

2.12 (2.3 节) 证明: 如果对于分布

$$N=\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}\right],$$

有 $\Pr\{X\geqslant 0,Y\geqslant 0\}=\alpha$, 则 $\rho=\cos(1-2\alpha)\pi$. [提示: 令 $X=U,Y=\rho U+\sqrt{1-\rho^2}V$, 并用几何方法证明 $\rho=\cos 2\pi(\frac{1}{2}-\alpha)$.]

2.13 (2.3 节) 证明: 如果 $\rho_{ij}=\rho$, $i\neq j$, $i,j=1,\cdots,p$, 则 $\rho\geqslant -1/(p-1)$.

2.14 (2.3 节) 同心椭球. 设 p 维向量 $\boldsymbol{Y}$, 当 $\boldsymbol{y}'\boldsymbol{y}\leqslant p+2$ 时的密度为 $f(\boldsymbol{y})=\Gamma(\frac{1}{2}p+1)/[(p+2)\pi]^{\frac{1}{2}p}$, 否则为 0. 则 $\mathrm{E}(\boldsymbol{Y})=0$ 且 $\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')=\boldsymbol{I}$ (习题 7.4). 由此结果证明: 如果 $\boldsymbol{X}$ 的密度当 $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})\leqslant p+2$ 时是 $g(\boldsymbol{x})=\sqrt{|\boldsymbol{A}|}\Gamma(\frac{1}{2}p+1)/[(p+2)\pi]^{\frac{1}{2}p}$, 否则为 0, 则 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{\mu}$ 且 $\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})']=\boldsymbol{A}^{-1}$.

2.15 (2.4 节) 证明当 $\boldsymbol{X}$ 为正态分布时, 其分量相互独立当且仅当协方差阵是对角的.

2.16 (2.4 节) 求关于 $\boldsymbol{A}$ 使得 $\boldsymbol{A}\boldsymbol{Y}+\boldsymbol{\lambda}$ 有连续的 cdf 的充分必要条件.

2.17 (2.4 节) 习题 2.7 中哪些密度决定的分布使得 X 和 Y 独立?

2.18 (2.4 节)

(a) 写出习题 2.6 中每个情况下 X 的边缘密度.

(b) 用记号 $N(a,b)$ 表示习题 2.7 中每个情况下 X 的边缘分布.

(c) 写出习题 2.9 中 X_1 和 X_2 的边缘密度.

2.19 (2.4 节) 当 X 和 Y 分别有习题 2.6 中的每个密度时, $Z=X-Y$ 的分布是什么?

2.20 (2.4 节) 当 X_1,X_2,X_3 服从习题 2.9 中的分布时, $X_1+2X_2-3X_3$ 的分布是什么?

2.21 (2.4 节) 设 $\boldsymbol{X}=(X_1,X_2)'$, 其中 $X_1=X$, $X_2=aX+b$, X 有分布 $N(0,1)$. 求 $\boldsymbol{X}$

的 cdf.

2.22 (2.4 节) 设 $X_1, \cdots, X_N$ 是独立分布的, 每一个都服从 $N(\mu, \sigma^2)$.

(a) $\boldsymbol{X} = (X_1, \cdots, X_N)'$ 的分布是什么? 求其均值向量和协方差阵.

(b) 用定理 2.4.4, 求 $\bar{X} = \sum X_i/N$ 的边缘分布.

2.23 (2.4 节) 设 $X_1, \cdots, X_N$ 是独立分布的, X_i 服从分布 $N(\beta + \gamma z_i, \sigma^2)$, 其中 z_i 是给定的数, $i = 1, \cdots, N$ 且 $\sum_i z_i = 0$.

(a) 求 $\boldsymbol{X} = (X_1, \cdots, X_N)'$ 的分布.

(b) 求 $\bar{X}$ 的分布和 $g = \sum X_i z_i / \sum z_i^2$ 的分布, 其中 $\sum z_i^2 > 0$.

2.24 (2.4 节) 设 $(X_1, Y_1)', (X_2, Y_2)', (X_3, Y_3)'$ 是独立分布的. $(X_i, Y_i)'$ 服从分布

$$N = \left[\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}\right], \qquad i = 1, 2, 3.$$

(a)求这六个变量的分布. (b) 求 $(\bar{X}, \bar{Y})'$ 的分布.

2.25 (2.4 节) 设 $\boldsymbol{X}$ 服从 (奇异) 正态分布, 其均值向量为 $\mathbf{0}$, 协方差阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}.$$

(a) 证明 $\boldsymbol{\Sigma}$ 的秩为 1.

(b) 求 $\boldsymbol{a}$, 使得 $\boldsymbol{X} = \boldsymbol{a}'Y$ 且 Y 服从非奇异正态分布, 并给出 Y 的密度.

2.26 (2.4 节) 设

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & -1 & 3 \\ -1 & 5 & -3 \\ 3 & -3 & 5 \end{pmatrix}.$$

(a) 求向量 $\boldsymbol{\mu} \neq \mathbf{0}$ 使得 $\boldsymbol{\Sigma\mu} = \mathbf{0}$. [提示: 对任意列取余子式.]

(b) 证明: 任意形如 $\boldsymbol{G} = (\boldsymbol{H} \;\; \boldsymbol{\mu})$($\boldsymbol{H}$ 是 3×2 的) 的矩阵有性质

$$\boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} = \begin{pmatrix} \boldsymbol{H}'\boldsymbol{\Sigma}\boldsymbol{H} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}.$$

(c) 利用 (a) 和 (b), 求满足 (36) 的 $\boldsymbol{B}$.

(d) 求 $\boldsymbol{B}^{-1}$ 和对应 (39) 的划分.

(e) 证明 $\boldsymbol{CC}' = \boldsymbol{\Sigma}$.

2.27 (2.4 节) 证明: 如果 X_1 和 X_2 的联合 (边缘) 分布是奇异的 (即退化的), 则 X_1, X_2 和 X_3 的联合分布是奇异的.

2.28 (2.5 节) 在习题 2.6 的每种情况下, 求给定 $Y = y$ 时 X 的条件分布和给定 $X = x$ 时 Y 的条件分布, 并在习题 2.6 中恰当的图上画出回归线.

2.29 (2.5 节) 设 $\boldsymbol{\mu} = \mathbf{0}$ 且

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.80 & -0.40 \\ 0.80 & 1 & -0.56 \\ -0.40 & -0.56 & 1 \end{pmatrix}.$$

(a) 求给定 $X_2 = x_2$ 时 X_1 和 X_3 的条件分布.

(b) 给定 X_2 时 X_1 和 X_3 的偏相关系数是多少?

2.30 (2.5 节) 在习题 2.9 中, 求给定 $X_3 = x_3$ 时 X_1 和 X_2 的条件分布.

2.31 (2.5 节) 直接用定理 2.5.1 证明 (20).

2.32 (2.5 节)

(a) 证明: 求 $\boldsymbol{\alpha}$ 使得极大化 X_i 和 $\boldsymbol{\alpha}'\boldsymbol{X}^{(2)}$ 的相关系数的绝对值, 等价于在 $\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{22}\boldsymbol{\alpha}$ 是常数的约束下极大化 $(\boldsymbol{\sigma}'_{(i)}\boldsymbol{\alpha})^2$.

(b) 求 $\boldsymbol{\alpha}$ 使得极大化 $(\boldsymbol{\sigma}'_{(i)}\boldsymbol{\alpha})^2 - \lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{22}\boldsymbol{\alpha} - c)$, 其中 c 是常数, λ 是拉格朗日乘子.

2.33 (2.5 节) 多重相关系数的不变性. 证明: $\bar{R}_{i\cdot q+1,\cdots,p}$ 是服从多元正态分布的 X_i 和 $\boldsymbol{X}^{(2)}$ 在变换 $x_i^* = b_i x_i + c_i$ $(b_i \neq 0)$ 和 $\boldsymbol{X}^{(2)*} = \boldsymbol{H}\boldsymbol{X}^{(2)} + \boldsymbol{k}$ ($\boldsymbol{H}$ 非奇异) 下的一个不变特征, 并且, 如果任何一个关于 $\mu_i, \sigma_{ii}, \boldsymbol{\sigma}_{(i)}, \boldsymbol{\mu}^{(2)}$ 和 $\boldsymbol{\Sigma}_{22}$ 的函数在此变换下是不变的, 则它一定是 $\bar{R}_{i\cdot q+1,\cdots,p}$ 的函数.

2.34 (2.5 节) 证明

$$1 - \bar{R}^2_{i\cdot q+1,\cdots,p} = \frac{1}{|\rho_{kj}|}\begin{vmatrix} 1 & \rho_{ij} \\ \rho_{ki} & \rho_{kj} \end{vmatrix}, \qquad k,j = q+1,\cdots,p.$$

2.35 (2.5 节) 求习题 2.29 中 X_1 和 (X_2, X_3) 的多重相关系数.

2.36 (2.5 节) 明确地证明: 如果 $\boldsymbol{\Sigma}$ 正定, 则

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}| \cdot |\boldsymbol{\Sigma}_{22}|.$$

2.37 (2.5 节) 证明阿达马不等式

$$|\boldsymbol{\Sigma}| \leqslant \prod_{i=1}^{p} \sigma_{ii}.$$

[提示: 利用习题 2.36, 证明 $|\boldsymbol{\Sigma}| \leqslant \sigma_{11}|\boldsymbol{\Sigma}_{22}|$, 其中 $\boldsymbol{\Sigma}_{22}$ 是 $(p-1)\times(p-1)$ 的, 并应用归纳法.]

2.38 (2.5 节) 证明习题 2.37 中的等号成立当且仅当 $\boldsymbol{\Sigma}$ 是对角的.

2.39 (2.5 节) 证明 $\beta_{12\cdot3} = \sigma_{12\cdot3}/\sigma_{22\cdot3} = \rho_{13\cdot2}\sigma_{1\cdot2}/\sigma_{3\cdot2}$ 且 $\beta_{13\cdot2} = \sigma_{13\cdot2}/\sigma_{33\cdot2} = \rho_{13\cdot2}\sigma_{1\cdot2}/\sigma_{3\cdot2}$, 其中 $\sigma^2_{i\cdot k} = \sigma_{ii\cdot k}$.

2.40 (2.5 节) 设 (X_1, X_2) 有密度 $n(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma}) = f(x_1, x_2)$. 设给定 $X_1 = x_1$ 时 X_2 的密度是 $f(x_2|x_1)$. 设 X_1, X_2, X_3 的联合密度是 $f(x_1, x_2)f(x_3|x_1)$. 求 X_1, X_2, X_3 的协方差阵和给定 X_1 时, X_2 和 X_3 的偏相关系数.

2.41 (2.5 节) 证明 $1 - \bar{R}^2_{1\cdot23} = (1-\rho^2_{13})(1-\rho^2_{12\cdot3})$. [提示: 利用 X_1 在给定 x_2 和 x_3 下的条件分布的方差是 $(1-\bar{R}^2_{1\cdot23})\sigma_{11}$.]

2.42 (2.5 节) 如果 $p=2$, 则 X_1 和 x_2 的简单相关系数与 X_1 和 $\boldsymbol{X}^{(2)} = X_2$ 的多重相关系数有差别么?

2.43 (2.5 节) 证明

$$\beta_{ik\cdot q+1,\cdots,k-1,k+1,\cdots,p} = \frac{\sigma_{ik\cdot q-1,\cdots,k-1,k+1,\cdots,p}}{\sigma_{kk\cdot q+1,\cdots,k-1,k+1,\cdots,p}} = \rho_{ik\cdot q-1,\cdots,k-1,k+1,\cdots,p}\frac{\sigma_{i\cdot q+1,\cdots,k-1,k+1,\cdots,p}}{\sigma_{k\cdot q+1,\cdots,k-1,k+1,\cdots,p}},$$

$i = 1,\cdots,q$, $k = q+1,\cdots,p$, 其中 $\sigma^2_{j\cdot q+1,\cdots,k-1,k+1,\cdots,p} = \sigma_{jj\cdot q+1,\cdots,k-1,k+1,\cdots,p}$, $j = i,k$. [提示: 利用习题 2.56, 取 $p_1 = q, p_2 = 1, p_3 = p-q-1$, 证明特殊情况 $k = q+1$.]

2.44 (2.5 节) 以 $\sigma_{i,q+1},\cdots,\sigma_{ip}$ 的角度给出 $\bar{R}_{i\cdot q+1,\cdots,p}=0$ 的充要条件.

2.45 (2.5 节) 证明

$$1-\bar{R}^2_{i\cdot q+1,\cdots,p}=(1-\rho^2_{ip})(1-\rho^2_{i,p-1\cdot p})\cdots(1-\rho^2_{i,q-1\cdot q+2,\cdots,p}).$$

[提示: 相继利用 (19) 和 (27).]

2.46 (2.5 节) 证明

$$\rho^2_{ij\cdot q+1,\cdots,p}=\beta_{ij\cdot q+1,\cdots,p}\beta_{ji\cdot q+1,\cdots,p}.$$

2.47 (2.5 节) 证明

$$\rho_{12\cdot 3\cdots p}=\frac{-\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}}.$$

[提示: 应用附录中的定理 A.3.2 计算 σ^{ij} 的余子式.]

2.48 (2.5 节) 证明: 对于任何期望存在的联合分布和任何函数 $h(\boldsymbol{x}^{(2)})$, 有

$$\mathrm{E}[(X_i-\mathrm{E}(X_i)|\boldsymbol{X}^{(2)})h(\boldsymbol{X}^{(2)})]=0.$$

[提示: 对于上式, 先取关于 X_i 在条件 $\boldsymbol{X}^{(2)}$ 下的期望.]

2.49 (2.5 节) 证明: 对于任意函数 $h(\boldsymbol{x}^{(2)})$ 和期望存在的 X_i 和 $\boldsymbol{X}^{(2)}$ 的联合分布, $\mathrm{E}[X_i-h(\boldsymbol{X}^{(2)})]^2=\mathrm{E}[X_i-g(\boldsymbol{X}^{(2)})]^2+\mathrm{E}[g(\boldsymbol{X}^{(2)})-h(\boldsymbol{X}^{(2)})]^2$, 其中 $g(\boldsymbol{x}^{(2)})=\mathrm{E}(X_i|\boldsymbol{x}^{(2)})$ 是给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 X_i 的条件期望. 因此, $g(\boldsymbol{X}^{(2)})$ 极小化预测的均方误差. [提示: 利用习题 2.48.]

2.50 (2.5 节) 证明: 对任意函数 $h(\boldsymbol{x}^{(2)})$ 和期望存在的 X_i 和 $\boldsymbol{X}^{(2)}$ 的联合分布, X_i 和 $h(\boldsymbol{X}^{(2)})$ 的相关系数不超过 X_i 和 $g(\boldsymbol{X}^{(2)})$ 的相关系数, 其中 $g(\boldsymbol{x}^{(2)})=\mathrm{E}(X_i|\boldsymbol{x}^{(2)})$.

2.51 (2.5 节) 证明: 对任意向量函数 $h(\boldsymbol{x}^{(2)})$,

$$\mathrm{E}[\boldsymbol{X}^{(1)}-h(\boldsymbol{X}^{(2)})][\boldsymbol{X}^{(1)}-h(\boldsymbol{X}^{(2)})]'-\mathrm{E}[\boldsymbol{X}^{(1)}-\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)})][\boldsymbol{X}^{(1)}-\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)})]'$$

是半正定的. 注意, 此结果推广了定理 2.5.3 和习题 2.49.

2.52 (2.5 节) 证明 $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}=-\boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}$, 其中 $\boldsymbol{\Psi}=\boldsymbol{\Sigma}^{-1}$ 有相应于 $\boldsymbol{\Sigma}$ 的分块.

2.53 (2.5 节) 证明

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}&=\begin{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2}^{-1} & -\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11\cdot 2}^{-1} & \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}+\boldsymbol{\Sigma}_{22}^{-1}\end{pmatrix}\\ &=\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1}\end{pmatrix}+\begin{pmatrix}\boldsymbol{I}\\ -\boldsymbol{\beta}'\end{pmatrix}\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{I}-\boldsymbol{\beta}),\end{aligned}$$

其中 $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$. [提示: 利用附录中的定理 A.3.3 和 $\boldsymbol{\Sigma}^{-1}$ 对称的事实.]

2.54 (2.5 节) 利用习题 2.53 证明

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}=(\boldsymbol{x}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}^{(2)})'\boldsymbol{\Sigma}_{11\cdot 2}^{-1}(\boldsymbol{x}^{(1)}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}^{(2)})+\boldsymbol{x}^{(2)\prime}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{x}^{(2)}.$$

2.55 (2.5 节) 证明

$$\begin{aligned}\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)})=&\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}(\boldsymbol{x}^{(3)}-\boldsymbol{\mu}^{(3)})\\ &+(\boldsymbol{\Sigma}_{12}-\boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32})(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32})^{-1}\\ &\cdot[\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}(\boldsymbol{x}^{(3)}-\boldsymbol{\mu}^{(3)})].\end{aligned}$$

2.56 (2.5 节) 用矩阵代数证明

$$\boldsymbol{\Sigma}_{11}-(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{13})\begin{pmatrix}\boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23}\\ \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\Sigma}_{21}\\ \boldsymbol{\Sigma}_{31}\end{pmatrix}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31}$$
$$-(\boldsymbol{\Sigma}_{12}-\boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32})(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{32})^{-1}(\boldsymbol{\Sigma}_{21}-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{31}).$$

2.57 (2.5 节) **偏相关系数的不变性**. 证明 $\rho_{12\cdot3,\cdots,p}$ 在变换 $x_i^*=a_ix_i+\boldsymbol{b}_i'\boldsymbol{x}^{(3)}+c_i(a_i>0,i=1,2,\ \boldsymbol{x}^{(3)*}=\boldsymbol{C}\boldsymbol{x}^{(3)}+\boldsymbol{d}$, 其中 $\boldsymbol{x}^{(3)}=(x_3,\cdots,x_p)')$ 下是不变的. 且任何在这些变换下不变的 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的函数是 $\rho_{12\cdot3,\cdots,p}$ 的函数.

2.58 (2.5 节) 假设 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 分别是 q 维和 $p-q$ 维向量, 有密度

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}}\mathrm{e}^{-\frac{1}{2}Q},$$

其中

$$Q=(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{A}_{11}(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})+(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{A}_{12}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})$$
$$+(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{A}_{21}(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})+(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{A}_{22}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}).$$

证明 Q 可以写成 Q_1+Q_2, 其中

$$Q_1=[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})]'\boldsymbol{A}_{11}[(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)})+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})],$$
$$Q_2=(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)})'(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}).$$

证明 $\boldsymbol{X}^{(2)}$ 的边缘密度是

$$\frac{|\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}(p-q)}}\mathrm{e}^{-\frac{1}{2}Q_2}.$$

证明给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件密度是

$$\frac{|\boldsymbol{A}_{11}|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}q}}\mathrm{e}^{-\frac{1}{2}Q_1}$$

(不用附录中的结论证明). 这个问题给出了定理 2.4.3 和 定理 2.5.1 的一个替换证明.

2.59 (2.6 节) 详细证明推论 2.6.2.

2.60 (2.6 节) 设 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. 对其特征函数微分, 证明 (25) 和 (26).

2.61 (2.6 节) 利用变换 $\boldsymbol{X}-\boldsymbol{\mu}=\boldsymbol{C}\boldsymbol{Y}$, 其中 $\boldsymbol{\sigma}=\boldsymbol{C}\boldsymbol{C}'$, 并利用对 $\boldsymbol{Y}$ 的密度积分, 证明 (25) 和 (26).

2.62 (2.6 节) 设 (X,Y) 的密度是

$$f(x)=\begin{cases}2n(x|0,1)n(y|0,1), & 0\leqslant y\leqslant x<\infty,0\leqslant -x\leqslant y<\infty,\\ & 0\leqslant -y\leqslant -x<\infty,0\leqslant x\leqslant -y<\infty,\\ 0, & \text{否则}.\end{cases}\tag{44}$$

证明 $X,Y,X+Y,X-Y$ 中每一个的边缘分布都是正态的.

2.63 (2.6 节) 假定 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. 设 $\boldsymbol{\Sigma}=(\boldsymbol{\sigma}_1,\cdots,\boldsymbol{\sigma}_p)$. 证明

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{X}'\otimes\boldsymbol{X}\boldsymbol{X}')=\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})'+\begin{bmatrix}\boldsymbol{\sigma}_1\boldsymbol{\sigma}_1' & \cdots & \boldsymbol{\sigma}_p\boldsymbol{\sigma}_1'\\ \vdots & & \vdots\\ \boldsymbol{\sigma}_1\boldsymbol{\sigma}_p' & \cdots & \boldsymbol{\sigma}_p\boldsymbol{\sigma}_p'\end{bmatrix}$$
$$=(\boldsymbol{I}+\boldsymbol{K})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})',$$

其中,

$$\mathrm{vec}\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_p \end{bmatrix}, \quad \boldsymbol{K} = \begin{bmatrix} \varepsilon_1\varepsilon_1' & \cdots & \varepsilon_p\varepsilon_1' \\ \vdots & & \vdots \\ \varepsilon_1\varepsilon_p' & \cdots & \varepsilon_p\varepsilon_p' \end{bmatrix},$$

且 ε_i 是第 i 个分量为 1、其他分量为 0 的列向量.

2.64 **复正态分布**. 设 $(\boldsymbol{X}', \boldsymbol{Y}')'$ 服从正态分布, 其均值向量为 $(\boldsymbol{\mu}_{\boldsymbol{X}}', \boldsymbol{\mu}_{\boldsymbol{Y}}')'$, 协方差阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma} & -\boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{\Gamma} \end{pmatrix},$$

其中, $\boldsymbol{\Gamma}$ 是正定的, 且 $\boldsymbol{\Phi} = -\boldsymbol{\Phi}'$(反对称). 则称 $\boldsymbol{Z} = \boldsymbol{X} + \mathrm{i}\boldsymbol{Y}$ 服从复正态分布, 其均值为 $\boldsymbol{\theta} = \boldsymbol{\mu}_{\boldsymbol{X}} + \mathrm{i}\boldsymbol{\mu}_{\boldsymbol{Y}}$, 协方差阵为 $\mathrm{E}(\boldsymbol{Z}-\boldsymbol{\theta})(\boldsymbol{Z}-\boldsymbol{\theta})^* = \boldsymbol{P} = \boldsymbol{Q} + \mathrm{i}\boldsymbol{R}$, 其中 $\boldsymbol{Z}^* = \boldsymbol{X}' - \mathrm{i}\boldsymbol{Y}'$. 注意 $\boldsymbol{P}$ 是埃尔米特正定矩阵.

(a) 证明 $\boldsymbol{Q} = 2\boldsymbol{\Gamma}$ 且 $\boldsymbol{R} = 2\boldsymbol{\Phi}$.

(b) 证明 $|\boldsymbol{P}|^2 = |2\boldsymbol{\Sigma}|$. [提示: $|\boldsymbol{\Gamma} + \mathrm{i}\boldsymbol{\Phi}| = |\boldsymbol{\Gamma} - \mathrm{i}\boldsymbol{\Phi}|$.]

(c) 证明

$$\boldsymbol{P}^{-1} = (\boldsymbol{Q} + \boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R})^{-1} - \mathrm{i}\boldsymbol{Q}^{-1}\boldsymbol{R}(\boldsymbol{Q} + \boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R})^{-1}.$$

注意埃尔米特矩阵的逆还是埃尔米特矩阵.

(d) 证明 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的密度可以记为

$$\pi^{-p}|\boldsymbol{P}|^{-1}\mathrm{e}^{-(\boldsymbol{z}-\boldsymbol{\theta})^*\boldsymbol{P}^{-1}(\boldsymbol{z}-\boldsymbol{\theta})}.$$

2.65 **复正态** (续). 如果 $\boldsymbol{Z}$ 服从习题 2.64 中的复正态分布, 证明 $\boldsymbol{W} = \boldsymbol{A}\boldsymbol{Z}$ 服从均值为 $\boldsymbol{A}\boldsymbol{\theta}$、协方差阵为 $\mathrm{C}(\boldsymbol{W}) = \boldsymbol{A}\boldsymbol{P}\boldsymbol{A}^*$ 的复正态分布, 其中 $\boldsymbol{A}$ 是非奇异复矩阵.

2.66 证明习题 2.64 中定义的 $\boldsymbol{Z}$ 的特征函数是

$$\mathrm{E}(\mathrm{e}^{\mathrm{i}R(\boldsymbol{\mu}^*\boldsymbol{Z})}) = \mathrm{e}^{\mathrm{i}R\boldsymbol{\mu}^*\boldsymbol{\theta} - \boldsymbol{\mu}^*\boldsymbol{P}\boldsymbol{\mu}},$$

其中 $\mathrm{R}(x + \mathrm{i}y) = x$.

2.67 (2.2 节) 证明 $\int_{-a}^{a} \mathrm{e}^{-x^2/2}\mathrm{d}x/\sqrt{2\pi}$ 近似于 $(1 - \mathrm{e}^{-2a^2/\pi})^{1/2}$. [提示: $(\boldsymbol{X}, \boldsymbol{Y})$ 落在一个正方形的概率近似于 $(\boldsymbol{X}, \boldsymbol{Y})$ 落在一个逼近圆的概率 [Pólya (1949)].]

2.68 (2.7 节) 对于密度为 (41) 的多元 t 分布, 证明 $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, $\mathrm{Cov}(\boldsymbol{X}) = [m/(m-2)]\boldsymbol{\Lambda}$.

第 3 章　均值向量和协方差阵的估计

3.1　引　　言

多元正态分布是完全由其均值向量 $\boldsymbol{\mu}$ 和协方差阵 $\boldsymbol{\Sigma}$ 决定的. 第一个统计问题是基于观测样本怎样估计这些参数. 3.2 节将证明 $\boldsymbol{\mu}$ 的极大似然估计是样本均值, $\boldsymbol{\Sigma}$ 的极大似然估计等比于样本方差和协方差的矩阵. 样本方差是观测值与其样本均值的离差平方和除以样本数减 1, 样本协方差从叉积的角度说有相似定义. 样本协方差阵是 $\boldsymbol{\Sigma}$ 的一个无偏估计.

3.3 节给出了样本均值向量的分布, 并且介绍了当 $\boldsymbol{\Sigma}$ 已知时, $\boldsymbol{\mu}$ 等于某已知向量的假设检验. $\boldsymbol{\Sigma}$ 未知的情况将在第 5 章中讨论.

3.4 节给出了样本均值的一些理论性质, 并给出了先验正态分布下, 总体均值的贝叶斯估计. 3.5 节介绍了 Tames-Stein 估计, 并讨论了关于均方误差损失函数的样本均值的改进.

3.6 节讨论了椭球等高分布的均值向量和协方差阵的估计, 并给出这些估计的分布.

3.2　均值向量和协方差阵的极大似然估计

给定来自 p 元 (非退化) 正态分布的 (向量) 观测样本, 我们需要求这个分布的均值向量 $\boldsymbol{\mu}$ 和协方差阵 $\boldsymbol{\Sigma}$ 的估计. 我们将推导它们的极大似然估计.

对于多元正态分布, 极大似然方法在求各种估计和假设检验的问题上非常有用. 极大似然估计及其修正常有优良的性质. 对于这里研究的特殊情况, 这些估计量是渐近有效的 [Cramér (1946), 33.3 节].

假设 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 我们的样本是 N 个观测值 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, 其中 $N > p$. 它们的似然函数是

$$\begin{aligned} L &= \prod_{\alpha=1}^{N} n(\boldsymbol{x}_\alpha | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad (1) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}pN} |\boldsymbol{\Sigma}|^{\frac{1}{2}N}} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right]. \end{aligned}$$

在似然函数中, 向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是由样本值给定的, 故 L 是 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的一个函数.

为了强调这些量是变量 (而不是参数), 我们用 $\boldsymbol{\mu}^*$ 和 $\boldsymbol{\Sigma}^*$ 表示它们. 则这个似然函数的对数是

$$\begin{aligned}\ln L = & -\frac{1}{2}pN\ln 2\pi - \frac{1}{2}N\ln|\boldsymbol{\Sigma}^*| \\ & -\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}^*).\end{aligned} \tag{2}$$

因为 $\ln L$ 是 L 的增函数, 所以它的极大值点也是 $\boldsymbol{\mu}^*$, $\boldsymbol{\Sigma}^*$ 空间里使得 L 极大的点. $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ 的极大似然估计就是极大化 $\ln L$ 的向量 $\boldsymbol{\mu}^*$ 和正定矩阵 $\boldsymbol{\Sigma}^*$. (下面考虑 $\ln L$ 关于正定矩阵 $\boldsymbol{\Sigma}^*$ 的上确界.)

设样本均值向量是

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha = \begin{pmatrix} \frac{1}{N}\sum_{\alpha=1}^{N} x_{1\alpha} \\ \vdots \\ \frac{1}{N}\sum_{\alpha=1}^{N} x_{p\alpha} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}, \tag{3}$$

其中 $\boldsymbol{x}_\alpha = (x_{1\alpha}, \cdots, x_{p\alpha})'$, $\bar{x}_i = \sum_{\alpha=1}^{N} x_{i\alpha}/N$, 且令关于均值的离差的叉积及平方和矩阵为

$$\begin{aligned}\boldsymbol{A} & = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' \\ & = \left[\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)\right], \qquad i, j = 1, \cdots, p.\end{aligned} \tag{4}$$

使用下面的引理会带来方便.

引理 3.2.1　设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是 N 个 (p 维) 向量, 且 $\bar{\boldsymbol{x}}$ 的定义如 (3). 则对于任意向量 $\boldsymbol{b}$, 有

$$\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{b})(\boldsymbol{x}_\alpha - \boldsymbol{b})' = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' + N(\bar{\boldsymbol{x}} - \boldsymbol{b})(\bar{\boldsymbol{x}} - \boldsymbol{b})'. \tag{5}$$

证明

$$\begin{aligned}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{b})(\boldsymbol{x}_\alpha - \boldsymbol{b})' & = \sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}) + (\bar{\boldsymbol{x}} - \boldsymbol{b})][(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}) + (\bar{\boldsymbol{x}} - \boldsymbol{b})]' \\ & = \sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' + (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}} - \boldsymbol{b})' \\ & \quad + (\bar{\boldsymbol{x}} - \boldsymbol{b})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' + (\boldsymbol{x} - \boldsymbol{x}\boldsymbol{b})(\bar{\boldsymbol{x}} - \boldsymbol{b})'] \\ & = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' + \left[\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})\right](\bar{\boldsymbol{x}} - \boldsymbol{b})' \\ & \quad + (\bar{\boldsymbol{x}} - \boldsymbol{b})\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' + N(\bar{\boldsymbol{x}} - \boldsymbol{b})(\bar{\boldsymbol{x}} - \boldsymbol{b})'.\end{aligned} \tag{6}$$

右端第二项和第三项是 0, 这是因为由 (3) 知 $\sum(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})=\sum\boldsymbol{x}_\alpha-N\bar{\boldsymbol{x}}=\mathbf{0}$. ■

当我们令 $\boldsymbol{b}=\boldsymbol{\mu}^*$ 时, 有

$$\begin{aligned}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)'&=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)' \\ &=\boldsymbol{A}+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'.\end{aligned}\tag{7}$$

利用这个结论和矩阵的迹的性质 ($\mathrm{tr}\boldsymbol{CD}=\sum c_{ij}d_{ji}=\mathrm{tr}\boldsymbol{DC}$), 我们得到

$$\begin{aligned}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)&=\mathrm{tr}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*) \\ &=\mathrm{tr}\sum_{\alpha=1}^{N}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)' \\ &=\mathrm{tr}\boldsymbol{\Sigma}^{*-1}\boldsymbol{A}+\mathrm{tr}\boldsymbol{\Sigma}^{*-1}N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)' \\ &=\mathrm{tr}\boldsymbol{\Sigma}^{*-1}\boldsymbol{A}+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*).\end{aligned}\tag{8}$$

因此, 我们可将 (2) 记为

$$\begin{aligned}\ln L=&-\frac{1}{2}pN\ln(2\pi)-\frac{1}{2}N\ln|\boldsymbol{\Sigma}^*| \\ &-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{*-1}\boldsymbol{A}-\frac{1}{2}N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*).\end{aligned}\tag{9}$$

因为 $\boldsymbol{\Sigma}^*$ 是正定的, 所以 $\boldsymbol{\Sigma}^{*-1}$ 也是正定的, 于是 $N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)\geqslant 0$, 并且当且仅当 $\boldsymbol{\mu}^*=\bar{\boldsymbol{x}}$ 时等于 0. 为了极大化 (9) 的第二项和第三项, 我们利用下面的引理 (后面的章节也将用到这个引理).

引理 3.2.2 *如果 $\boldsymbol{D}$ 是 p 阶正定矩阵, 那么*

$$f(\boldsymbol{G})=N\ln|\boldsymbol{G}|-\mathrm{tr}\boldsymbol{G}^{-1}\boldsymbol{D}\tag{10}$$

关于正定矩阵 $\boldsymbol{G}$ 的极大值存在, 极大值出现在 $\boldsymbol{G}=(1/N)\boldsymbol{D}$ 处, 且为

$$f[(1/N)\boldsymbol{D}]=pN\ln N-N\ln|\boldsymbol{D}|-pN.\tag{11}$$

证明 设 $\boldsymbol{D}=\boldsymbol{E}\boldsymbol{E}'$ 且 $\boldsymbol{E}'\boldsymbol{G}^{-1}\boldsymbol{E}=\boldsymbol{H}$. 则 $\boldsymbol{G}=\boldsymbol{E}\boldsymbol{H}^{-1}\boldsymbol{E}'$, 从而 $|\boldsymbol{G}|=|\boldsymbol{E}|\cdot|\boldsymbol{H}^{-1}|\cdot|\boldsymbol{E}'|=|\boldsymbol{H}^{-1}|\cdot|\boldsymbol{E}\boldsymbol{E}'|=|\boldsymbol{D}|/|\boldsymbol{H}|$ 且 $\mathrm{tr}\boldsymbol{G}^{-1}\boldsymbol{D}=\mathrm{tr}\boldsymbol{G}^{-1}\boldsymbol{E}\boldsymbol{E}'=\mathrm{tr}\boldsymbol{E}'\boldsymbol{G}^{-1}\boldsymbol{E}=\mathrm{tr}\boldsymbol{H}$. 则要被极大化的 (关于正定矩阵 $\boldsymbol{H}$) 函数是

$$f=-N\ln|\boldsymbol{D}|+N\ln|\boldsymbol{H}|-\mathrm{tr}\boldsymbol{H}.\tag{12}$$

设 $\boldsymbol{H}=\boldsymbol{T}\boldsymbol{T}'$, 其中 $\boldsymbol{T}$ 是下三角的 (推论 A.1.7). 则

$$\begin{aligned}f&=-N\ln|\boldsymbol{D}|+N\ln|\boldsymbol{T}|^2-\mathrm{tr}\boldsymbol{T}\boldsymbol{T}' \\ &=-N\ln|\boldsymbol{D}|+\sum_{i=1}^{p}(N\ln t_{ii}^2-t_{ii}^2)-\sum_{i>j}t_{ij}^2\end{aligned}\tag{13}$$

的极大值出现在 $t_{ii}^2=N, t_{ij}=0, i\neq j$, 即 $\boldsymbol{H}=N\boldsymbol{I}$. 于是 $\boldsymbol{G}=(1/N)\boldsymbol{E}\boldsymbol{E}'=(1/N)\boldsymbol{D}$. ■

定理 3.2.1 如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu}\boldsymbol{\Sigma})$ $(p < N)$ 的样本, 则 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的极大似然估计分别是 $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = (1/N)\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha$ 和 $\hat{\boldsymbol{\Sigma}} = (1/N)\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$.

Anderson and Olkin (1985) 讨论了其他获得极大似然估计的方法. 见习题 3.4、习题 3.8 和习题 3.12.

通过特殊化引理 3.2.1 ($\boldsymbol{b}$=$\boldsymbol{0}$), 估计 $\hat{\boldsymbol{\Sigma}}$ 的计算会更简单,

$$\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'. \tag{14}$$

$\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha'$ 的元素是 $\sum_{\alpha=1}^{N} x_{i\alpha}x_{j\alpha}$, $N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'$ 的元素是 $N\bar{x}_i\bar{x}_j$ 或者 $(\sum_{\alpha=1}^{N} x_{i\alpha})(\sum_{\alpha=1}^{N} x_{j\alpha})/N$. 需要注意的是, 如果 $N > p$, 抽取样本使得 (14) 正定的概率是 1, 见习题 3.17.

协方差阵可以写成方差 (或标准差) 和相关系数的形式. 它们由方差和协方差唯一决定. 我们将说明, 参数的函数的极大似然估计是参数的极大似然估计的该函数.

引理 3.2.3 设 $f(\theta)$ 是集合 S 上的实值函数, ϕ 是从集合 S 到集合 S^* 的单值函数, 且有单值逆函数; 即对每一个 $\theta \in S$, 有唯一的 $\theta^* \in S^*$ 与之对应而且对每一个 $\theta^* \in S^*$, 有唯一的 $\theta \in S$ 与之对应. 设

$$g(\theta^*) = f[\phi^{-1}(\theta^*)]. \tag{15}$$

则如果 $f(\theta)$ 在 $\theta = \theta_0$ 达到极大值, 则 $g(\theta^*)$ 在 $\theta^* = \theta_0^* = \phi(\theta_0)$ 达到极大值. 如果 θ_0 是唯一使 $f(\theta)$ 达到极大值的点, 则 θ_0^* 也是唯一使 $g(\theta^*)$ 达到极大值的点.

证明 由假设, 对所有 $\theta \in S$, 有 $f(\theta_0) \geqslant f(\theta)$. 则对任意 $\theta^* \in S^*$, 有

$$g(\theta^*) = f[\phi^{-1}(\theta^*)] = f(\theta) \leqslant f(\theta_0) = g[\phi(\theta_0)] = g(\theta_0^*). \tag{16}$$

因此, $g(\theta^*)$ 在 θ^* 达到极大值. 如果 $f(\theta)$ 仅在 θ_0 达到极大值, 则对 $\theta \neq \theta_0$, 不等式严格成立, 从而 θ_0^* 也是 $g(\theta^*)$ 唯一的极大值点. ■

我们有下面的推论.

推论 3.2.1 对于给定的样本, 若 $\hat{\theta}_1, \cdots, \hat{\theta}_m$ 是某分布的参数 $\theta_1, \cdots, \theta_m$ 的极大似然估计, 则 $\phi_1(\hat{\theta}_1, \cdots, \hat{\theta}_m), \cdots, \phi_m(\hat{\theta}_1, \cdots, \hat{\theta}_m)$ 是 $\phi_1(\theta_1, \cdots, \theta_m), \cdots, \phi_m(\theta_1, \cdots, \theta_m)$ 的极大似然估计, 如果从 $\theta_1, \cdots, \theta_m$ 到 $\phi_1, \cdots, \phi_m$ 的变换是一一的.[①] 如果 $\theta_1, \cdots, \theta_m$ 的估计是唯一的, 那么 $\phi_1, \cdots, \phi_m$ 的估计也是唯一的.

推论 3.2.2 如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的样本, 其中 $\sigma_{ij} = \sigma_i\sigma_j\rho_{ij}$ $(\rho_{ii} = 1)$, 则 $\boldsymbol{\mu}$ 的极大似然估计是 $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = (1/N)\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha$; σ_i^2 的极大似然估计是 $\hat{\sigma}_i^2 = (1/N)\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2 = (1/N)(\sum_{\alpha=1}^{N} x_{i\alpha}^2 - N\bar{x}_i^2)$, 其中 $x_{i\alpha}$ 是 $\boldsymbol{x}_\alpha$ 的第 i 个分量, $\bar{x}_i$

① 变换是一一的假定条件是为了确保 $\phi_1 \cdots, \phi_m$ 唯一决定其似然. 若 $\theta^* = \phi(\theta)$ 没有唯一逆, 另一个选择是定义 $s(\theta^*) = \{\theta : \phi(\theta) = \theta^*\}$ 和 $g(\theta^*) = \sup f(\theta)|\theta \in S(\theta^*)$, 当 $f(\theta)$ 是似然函数时, $g(\theta^*)$ 称为"诱导似然". 则 $\hat{\theta}^* = \phi(\hat{\theta})$ 极大化 $g(\theta^*)$, 因为对所有 $\theta^* \in S^*$, $g(\theta^*) = \sup f(\theta)|\theta \in S(\theta^*) \geqslant \sup f(\theta)|\theta \in S = f(\hat{\theta}) = g(\hat{\theta}^*)$. [例如, 见 Zehna (1966).]

是 $\bar{x}$ 的第 i 个分量; ρ_{ij} 的极大似然估计是

$$\hat{\rho}_{ij} = \frac{\sum_{\alpha=1}^{N}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha}-\bar{x}_i)^2}\sqrt{\sum_{\alpha=1}^{N}(x_{j\alpha}-\bar{x}_j)^2}} = \frac{\sum_{\alpha=1}^{N}x_{i\alpha}x_{j\alpha}-N\bar{x}_i\bar{x}_j}{\sqrt{\sum_{\alpha=1}^{N}x_{i\alpha}^2-N\bar{x}_i^2}\sqrt{\sum_{\alpha=1}^{N}x_{j\alpha}^2-N\bar{x}_j^2}}. \tag{17}$$

证明 参数 $\mu_i=\mu_i, \sigma_i^2=\sigma_{ii}$ 和 $\rho_{ij}=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ 的集合是参数 μ_i 和 σ_{ij} 的集合的一一变换. 因此, 由推论 3.2.1 知, μ_i 的估计是 $\hat{\mu}_i$, σ_i^2 的估计是 $\hat{\sigma}_{ii}$, ρ_{ij} 的估计是

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}. \quad ■ \tag{18}$$

Pearson(1896) 给出了这个 ρ_{ij} 估计的一个解释, 因而有时称 (17) 为**皮尔逊相关系数**. 也称为**简单相关系数**. 常用 r_{ij} 表示.

一个关于样本 $(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)=\boldsymbol{X}$ 的方便的几何解释是从 $\boldsymbol{X}$ 的行的角度出发的. 设

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pN} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}_1' \\ \vdots \\ \boldsymbol{u}_p' \end{pmatrix}, \tag{19}$$

即 $\boldsymbol{u}_i'$ 是 $\boldsymbol{X}$ 的第 i 行. 向量 $\boldsymbol{u}_i$ 可以看作 N 维空间中一个端点的第 α 个坐标是 $x_{i\alpha}$、另一个端点在原点处的向量. 因此, 样本就被表示为 N 维欧氏空间里的 p 个向量. 由欧氏度量的定义, $\boldsymbol{u}_i$ 的平方长度 (即一个端点到另一个端点的平方距离) 是 $\boldsymbol{u}_i'\boldsymbol{u}_i=\sum_{\alpha=1}^{N}x_{i\alpha}^2$.

现在我们证明 $\boldsymbol{u}_i$ 和 $\boldsymbol{u}_j$ 的夹角的余弦是 $\boldsymbol{u}_i'\boldsymbol{u}_j/\sqrt{\boldsymbol{u}_i'\boldsymbol{u}_i\boldsymbol{u}_j'\boldsymbol{u}_j}=\sum_{\alpha=1}^{N}x_{i\alpha}x_{j\alpha}/\sqrt{\sum_{\alpha=1}^{N}x_{i\alpha}^2\sum_{\alpha=1}^{N}x_{j\alpha}^2}$. 取标量 d, 使得向量 $d\boldsymbol{u}_j$ 正交于 $\boldsymbol{u}_i-d\boldsymbol{u}_j$, 即 $0=d\boldsymbol{u}_j'(\boldsymbol{u}_i-d\boldsymbol{u}_j)=d(\boldsymbol{u}_j'\boldsymbol{u}_i-d\boldsymbol{u}_j'\boldsymbol{u}_j)$. 因此, $d=\boldsymbol{u}_j'\boldsymbol{u}_i/\boldsymbol{u}_j'\boldsymbol{u}_j$. 如图 3.1, 我们把 $\boldsymbol{u}_i$ 分解为 $\boldsymbol{u}_i-d\boldsymbol{u}_j$ 和 $d\boldsymbol{u}_j$ $[\boldsymbol{u}_i=(\boldsymbol{u}_i-d\boldsymbol{u}_j)+d\boldsymbol{u}_j]$. $\boldsymbol{u}_i$ 和 $\boldsymbol{u}_j$ 的夹角的余弦的绝对值是 $d\boldsymbol{u}_j$ 的长度除以 $\boldsymbol{u}_i$ 的长度, 即 $\sqrt{d\boldsymbol{u}_j'(d\boldsymbol{u}_j)/\boldsymbol{u}_i'\boldsymbol{u}_i}=\sqrt{d\boldsymbol{u}_j'\boldsymbol{u}_jd/\boldsymbol{u}_i'\boldsymbol{u}_i}$, 余弦是 $\boldsymbol{u}_i'\boldsymbol{u}_j/\sqrt{\boldsymbol{u}_i'\boldsymbol{u}_i\boldsymbol{u}_j'\boldsymbol{u}_j}$. 所要结果得证.

为了给出 a_{ii} 和 $a_{ij}/\sqrt{a_{ii}a_{jj}}$ 的几何解释, 我们引入等角线, 这条线穿过原点和点 $(1,1,\cdots,1)$. 如图 3.2. $\boldsymbol{u}_i$ 在向量 $\boldsymbol{\varepsilon}=(1,1,\cdots,1)'$ 上的投影是 $(\boldsymbol{\varepsilon}'\boldsymbol{u}_i/\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})\boldsymbol{\varepsilon}=(\sum_\alpha x_{i\alpha}/\sum_\alpha 1)\boldsymbol{\varepsilon}=\bar{x}_i\boldsymbol{\varepsilon}=(\bar{x}_i,\bar{x}_i\cdots,\bar{x}_i)'$. 则我们将 $\boldsymbol{u}_i$ 分解成 $\bar{x}_i\boldsymbol{\varepsilon}$(等角线上的投影) 和 $\boldsymbol{u}_i-\bar{x}_i\boldsymbol{\varepsilon}$($\boldsymbol{u}_i$ 在等角线的垂直线上的投影). $\boldsymbol{u}_i-\bar{x}_i\boldsymbol{\varepsilon}$ 的平方长度是 $(\boldsymbol{u}_i-\bar{x}_i\boldsymbol{\varepsilon})'(\boldsymbol{u}_i-\bar{x}_i\boldsymbol{\varepsilon})=\sum_\alpha(x_{i\alpha}-\bar{x}_i)^2$, 这就是 $N\hat{\sigma}=a_{ii}$. 平移 $\boldsymbol{u}_i-\bar{x}_i\boldsymbol{\varepsilon}$ 和 $\boldsymbol{u}_j-\bar{x}_j\boldsymbol{\varepsilon}$, 使得每个向量都有一个端点在原点; 第一个向量的第 α 个坐标是 $x_{i\alpha}-\bar{x}_i$, 第二个向量的第 α 个坐标是 $x_{j\alpha}-\bar{x}_j$, 两向量夹角的余弦是

$$r_{ij} = \frac{(\boldsymbol{u}_i - \bar{x}_i\boldsymbol{\varepsilon})'(\boldsymbol{u}_j - \bar{x}_j\boldsymbol{\varepsilon})}{(\sqrt{\boldsymbol{u}_i - \bar{x}_i\boldsymbol{\varepsilon})'(\boldsymbol{u}_i - \bar{x}_i\boldsymbol{\varepsilon})(\boldsymbol{u}_j - \bar{x}_j\boldsymbol{\varepsilon})'(\boldsymbol{u}_j - \bar{x}_j\boldsymbol{\varepsilon})}} \tag{20}$$

$$= \frac{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2 \sum_{\alpha=1}^{N}(x_{j\alpha} - \bar{x}_j)^2}}.$$

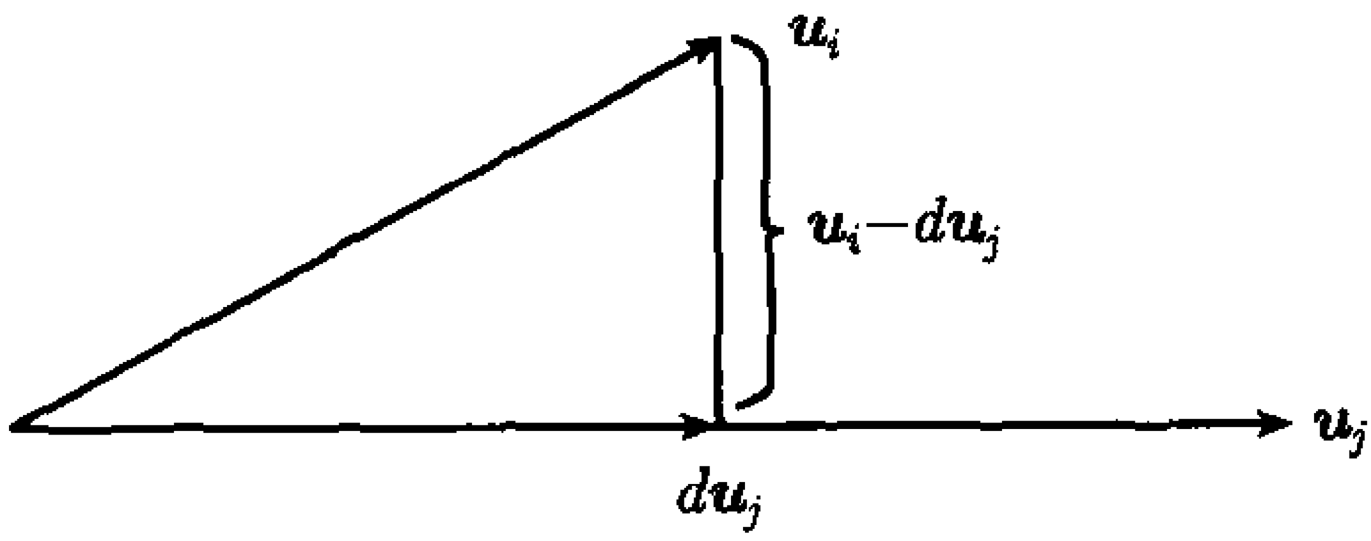


图 3.1

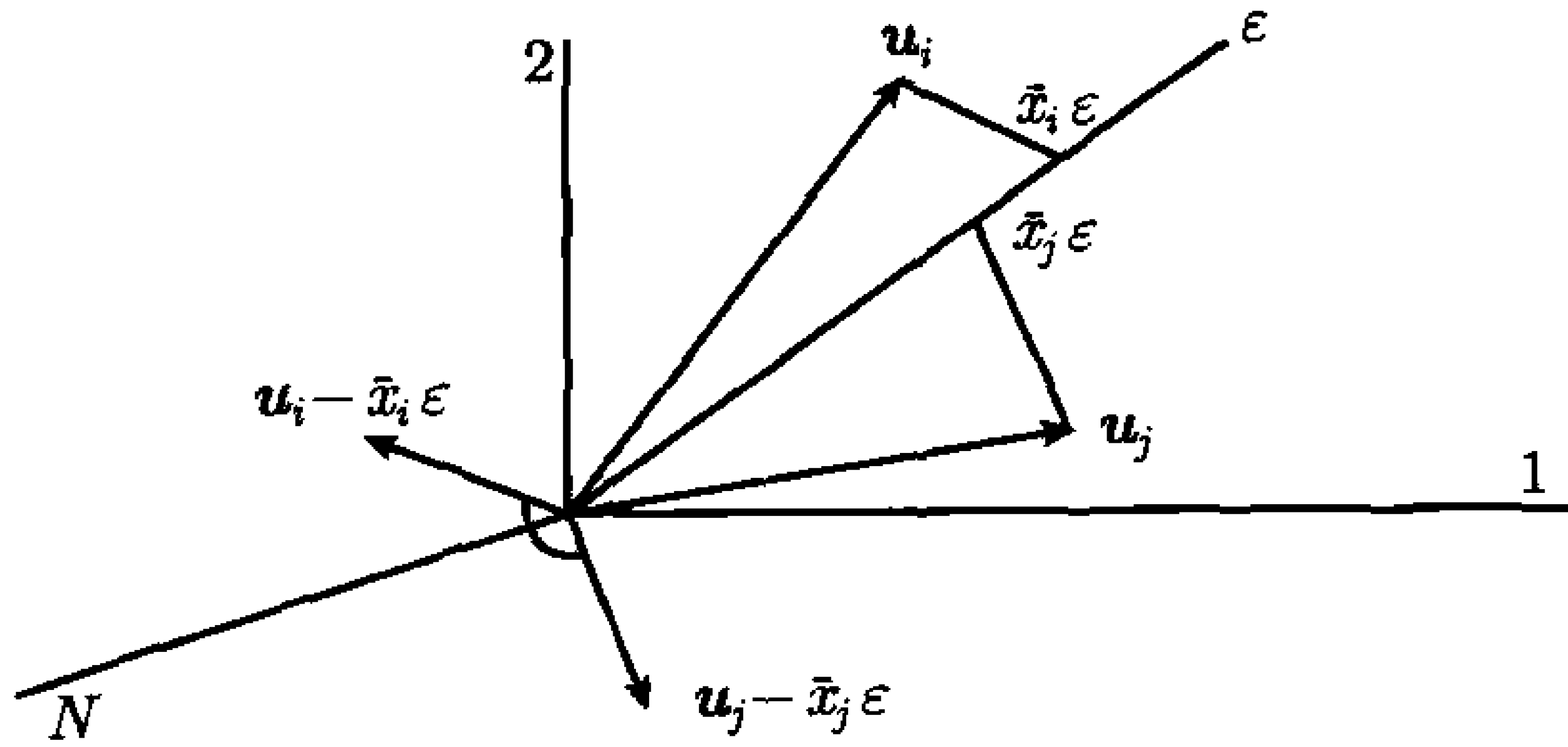


图 3.2

作为计算的一个例子, 考虑表 3.1 和图 3.3 表示的数据. 这些数据来自 Student (1908). 测量值 $x_{11} = 1.9$ 是第一个病人使用镇静剂 A 后, 睡眠时间增加的小时数, $x_{21} = 0.7$ 是该病人使用镇静剂 B 后, 睡眠时间增加的小时数, 等等. 假定每一对 (即表中每一行) 数据是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的一个观测, 我们求得

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \begin{pmatrix} 2.33 \\ 0.75 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 3.61 & 2.56 \\ 2.56 & 2.88 \end{pmatrix}, \tag{21}$$

$$\boldsymbol{S} = \begin{pmatrix} 4.01 & 2.85 \\ 2.85 & 3.20 \end{pmatrix},$$

$\hat{\rho}_{12} = r_{12} = 0.7952$. ($\boldsymbol{S}$ 将在后面定义.)

表 3.1 睡眠增长数据

病人	镇静剂 $\boldsymbol{A}$ x_1	镇静剂 $\boldsymbol{B}$ x_2
1	1.9	0.7
2	0.8	−1.6

(续)

病人	镇静剂 A x_1	镇静剂 B x_2
3	1.1	−0.2
4	0.1	−1.2
5	−0.1	−0.1
6	4.4	3.4
7	5.5	3.7
8	1.6	0.8
9	4.6	0.0
10	3.4	2.0

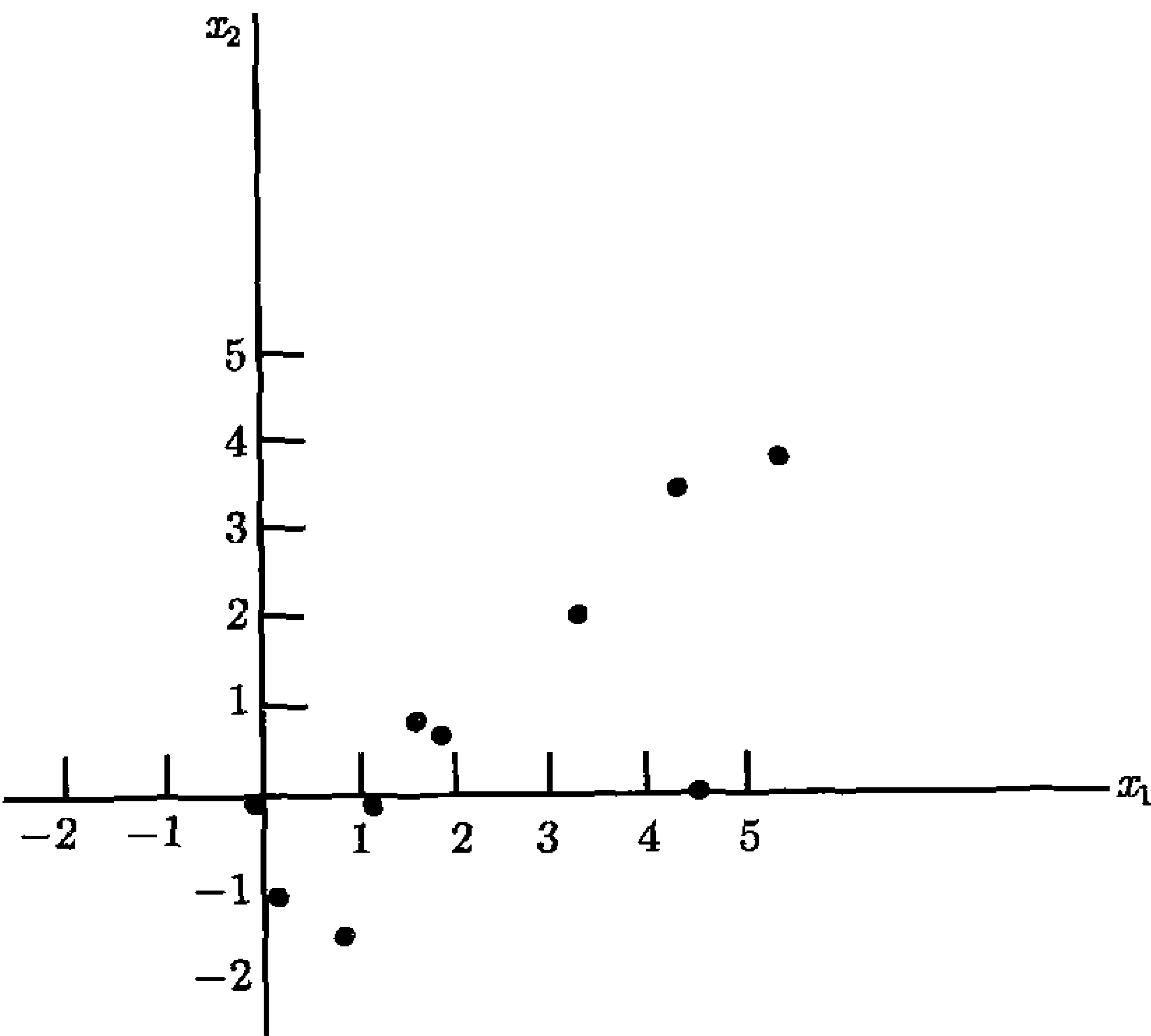


图 3.3 睡眠增长数据

3.3 样本均值向量的分布，协方差阵已知时均值的推断

3.3.1 分布理论

在单变量正态分布情形下，样本均值服从正态分布且和样本方差独立. 同样，3.2 节中定义的样本均值 $\bar{\boldsymbol{X}}$ 也是正态分布的，且和 $\hat{\boldsymbol{\Sigma}}$ 独立.

为了证明这个结果，我们将对观测向量集合做一个变换. 因为这种变换将在本书中出现多次，所以我们先证明一个更一般的定理.

定理 3.3.1 假定 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 独立，其中 $\boldsymbol{X}_\alpha$ 服从分布 $N(\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma})$. 设 $\boldsymbol{C}=(c_{\alpha\beta})$ 是一个 $N\times N$ 的正交矩阵. 则 $\boldsymbol{Y}_\alpha=\sum_{\beta=1}^N c_{\alpha\beta}\boldsymbol{X}_\beta$ 服从分布 $N(\boldsymbol{\nu}_\alpha,\boldsymbol{\Sigma})$，其中 $\boldsymbol{\nu}_\alpha=\sum_{\beta=1}^N c_{\alpha\beta}\boldsymbol{\mu}_\beta, \alpha=1,\cdots,N$，并且 $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_N$ 独立.

证明 向量 $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_N$ 具有联合正态分布，因为它们都是 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 的

分量的线性组合, 这些组合服从联合正态分布. $\boldsymbol{Y}_\alpha$ 的期望值是

$$\mathrm{E}(\boldsymbol{Y}_\alpha)=\mathrm{E}\left(\sum_{\beta=1}^{N} c_{\alpha\beta}\boldsymbol{X}_\beta\right)=\sum_{\beta=1}^{N} c_{\alpha\beta}\mathrm{E}(\boldsymbol{X}_\beta) \tag{1}$$
$$=\sum_{\beta=1}^{N} c_{\alpha\beta}\boldsymbol{\mu}_\beta=\boldsymbol{\nu}_\alpha.$$

$\boldsymbol{Y}_\alpha$ 与 $\boldsymbol{Y}_\gamma$ 的协方差阵是

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{Y}_\alpha,\boldsymbol{Y}'_\gamma)&=\mathrm{E}[(\boldsymbol{Y}_\alpha-\boldsymbol{\nu}_\alpha)(\boldsymbol{Y}_\gamma-\boldsymbol{\nu}_\gamma)'] \\ &=\mathrm{E}\left[\sum_{\beta=1}^{N} c_{\alpha\beta}(\boldsymbol{X}_\beta-\boldsymbol{\mu}_\beta)\right]\left[\sum_{\varepsilon=1}^{N} c_{\gamma\varepsilon}(\boldsymbol{X}_\varepsilon-\boldsymbol{\mu}_\varepsilon)'\right] \\ &=\sum_{\beta=1,\varepsilon=1}^{N} c_{\alpha\beta}c_{\gamma\varepsilon}\mathrm{E}[(\boldsymbol{X}_\beta-\boldsymbol{\mu}_\beta)(\boldsymbol{X}_\varepsilon-\boldsymbol{\mu}_\varepsilon)'] \\ &=\sum_{\beta,\varepsilon=1}^{N} c_{\alpha\beta}c_{\gamma\varepsilon}\delta_{\beta\varepsilon}\boldsymbol{\Sigma} \\ &=\sum_{\beta=1}^{N} c_{\alpha\beta}c_{\gamma\beta}\boldsymbol{\Sigma} \\ &=\delta_{\alpha\gamma}\boldsymbol{\Sigma},\end{aligned} \tag{2}$$

其中 $\delta_{\alpha\gamma}$ 是克罗内科符号 ($=1$ 如果 $\alpha=\gamma$, $=0$ 如果 $\alpha\neq\gamma$). 这说明 $\boldsymbol{Y}_\alpha$ 与 $\boldsymbol{Y}_\gamma$ 独立, $\alpha\neq\gamma$ 且 $\boldsymbol{Y}_\alpha$ 有协方差阵 $\boldsymbol{\Sigma}$. ■

我们也会用到下面的一般引理.

引理 3.3.1 *如果* $\boldsymbol{C}=(c_{\alpha\beta})$ *是正交的, 则* $\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{x}'_\alpha=\sum_{\alpha=1}^{N}\boldsymbol{y}_\alpha\boldsymbol{y}'_\alpha$, *其中* $\boldsymbol{y}_\alpha=\sum_{\beta=1}^{N}c_{\alpha\beta}\boldsymbol{x}_\beta,\alpha=1,\cdots,N$.

证明

$$\begin{aligned}\sum_{\alpha=1}^{N}\boldsymbol{y}_\alpha\boldsymbol{y}'_\alpha&=\sum_\alpha\sum_\beta c_{\alpha\beta}\boldsymbol{x}_\beta\sum_\gamma c_{\alpha\gamma}\boldsymbol{x}'_\gamma \\ &=\sum_{\beta,\gamma}\left(\sum_\alpha c_{\alpha\beta}c_{\alpha\gamma}\right)\boldsymbol{x}_\beta\boldsymbol{x}'_\gamma \\ &=\sum_{\beta,\gamma}\delta_{\beta\gamma}\boldsymbol{x}_\beta\boldsymbol{x}'_\gamma \\ &=\sum_\beta\boldsymbol{x}_\beta\boldsymbol{x}'_\beta.\end{aligned} \tag{3}$$

■

设 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 独立, 并且都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. 存在 $N\times N$ 正交矩阵 $\boldsymbol{B}=(b_{\alpha\beta})$, 其最后一行是

$$(1/\sqrt{N},\cdots,1/\sqrt{N}). \tag{4}$$

(见引理 A.4.2.) 这是 3.2 节描述的 N 维空间上的一个将等角线变为第 N 个坐标轴的旋转变换. 设 $\boldsymbol{A}=N\hat{\boldsymbol{\Sigma}}$, 如 3.2 节定义, 并设

$$\boldsymbol{Z}_\alpha=\sum_{\beta=1}^{N} b_{\alpha\beta}\boldsymbol{X}_\beta. \tag{5}$$

则

$$\boldsymbol{Z}_N=\sum_{\beta=1}^{N} b_{N\beta}\boldsymbol{X}_\beta=\sum_{\beta=1}^{N}\frac{1}{\sqrt{N}}\boldsymbol{X}_\beta=\sqrt{N}\bar{\boldsymbol{X}}. \tag{6}$$

由引理 3.3.1, 我们得到

$$\begin{aligned}\boldsymbol{A}&=\sum_{\alpha=1}^{N}\boldsymbol{X}_\alpha\boldsymbol{X}_\alpha'-N\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}'\\&=\sum_{\alpha=1}^{N}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'-\boldsymbol{Z}_N\boldsymbol{Z}_N'\\&=\sum_{\alpha=1}^{N-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'.\end{aligned} \tag{7}$$

因为 $\boldsymbol{Z}_N$ 与 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{N-1}$ 独立, 所以均值向量 $\bar{\boldsymbol{X}}$ 与 $\boldsymbol{A}$ 独立. 因为

$$\mathrm{E}(\boldsymbol{Z}_N)=\sum_{\beta=1}^{N} b_{N\beta}\mathrm{E}(\boldsymbol{X}_\beta)=\sum_{\beta=1}^{N}\frac{1}{\sqrt{N}}\boldsymbol{\mu}=\sqrt{N}\boldsymbol{\mu}, \tag{8}$$

所以 $\boldsymbol{Z}_N$ 服从分布 $N(\sqrt{N}\boldsymbol{\mu},\boldsymbol{\Sigma})$, 从而 $\bar{\boldsymbol{X}}=(1/\sqrt{N})\boldsymbol{Z}_N$ 服从分布 $N[\boldsymbol{\mu},(1/N)\boldsymbol{\Sigma}]$. 我们注意到

$$\begin{aligned}\mathrm{E}(\boldsymbol{Z}_\alpha)&=\sum_{\beta=1}^{N} b_{\alpha\beta}\mathrm{E}(\boldsymbol{X}_\beta)=\sum_{\beta=1}^{N} b_{\alpha\beta}\boldsymbol{\mu}\\&=\sum_{\beta=1}^{N} b_{\alpha\beta}b_{N\beta}\sqrt{N}\boldsymbol{\mu}\\&=0,\quad \alpha\neq N\end{aligned} \tag{9}$$

定理 3.3.2 来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 且容量为 N 的样本的均值服从分布 $N[\boldsymbol{\mu},(1/N)\boldsymbol{\Sigma}]$, 并且和 $\boldsymbol{\Sigma}$ 的极大似然估计 $\hat{\boldsymbol{\Sigma}}$ 独立. $N\hat{\boldsymbol{\Sigma}}$ 和 $\sum_{\alpha=1}^{N-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 同分布, 其中 $\boldsymbol{Z}_\alpha$ 服从分布 $N(\mathbf{0},\boldsymbol{\Sigma}),\alpha=1,\cdots,N-1$, 并且 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{N-1}$ 独立.

定义 3.3.1 一个参数向量 $\boldsymbol{\theta}$ 的估计 $\boldsymbol{t}$ 是无偏的当且仅当 $\mathrm{E}_{\boldsymbol{\theta}}\boldsymbol{t}=\boldsymbol{\theta}$.

因为 $\mathrm{E}(\bar{\boldsymbol{X}})=(1/N)\mathrm{E}(\sum_{\alpha=1}^{N}\boldsymbol{X}_\alpha)=\boldsymbol{\mu}$, 所以样本均值是总体均值的无偏估计. 但是

$$\mathrm{E}(\hat{\boldsymbol{\Sigma}})=\frac{1}{N}\mathrm{E}\left(\sum_{\alpha=1}^{N-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'\right)=\frac{N-1}{N}\boldsymbol{\Sigma}. \tag{10}$$

因此 $\hat{\boldsymbol{\Sigma}}$ 是 $\boldsymbol{\Sigma}$ 的有偏估计. 我们定义

$$S=\frac{1}{N-1}A=\frac{1}{N-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})' \tag{11}$$

为样本协方差阵. 它是 $\boldsymbol{\Sigma}$ 的一个无偏估计, 其对角线元素是 $\boldsymbol{X}$ 的分量的普通 (无偏) 样本方差.

3.3.2 协方差阵已知时均值向量的检验和置信区域

一个相当重要的统计问题是关于一个正态分布的均值向量等于某已知向量的假设检验及其相关问题, 即对未知均值向量给出置信区域. 现在我们在假定协方差阵 $\boldsymbol{\Sigma}$ 已知的情况下来研究这些问题. 我们在第 5 章中考虑协方差阵未知的情况.

在单变量情形中, 检验或置信区间基于样本均值与总体均值之差服从均值为零、方差已知的正态分布的事实, 从而可以用正态分布表设置分位数或计算置信区间. 在多元情形下, 可以利用样本均值向量和总体均值向量的差服从均值为 0 向量、协方差阵已知的正态分布的事实. 基于此分布, 可对每个分量设定限制, 但是它的缺点是限制的选择有些随意, 在对某些备择的检验问题中表现很差, 并且这些限制很难计算, 因为只有在二元情形下有表可用. 但是下面给出的一些方法更易于计算并有直观和理论的解释.

这些方法及其性质评价基于下面的定理.

定理 3.3.3 *如果 m 维向量 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{\nu},\boldsymbol{T})$ (非奇异), 则 $\boldsymbol{Y}'\boldsymbol{T}^{-1}\boldsymbol{Y}$ 服从自由度为 m、非中心参数为 $\boldsymbol{\nu}'\boldsymbol{T}^{-1}\boldsymbol{\nu}$ 的非中心 χ^2 分布. 如果 $\boldsymbol{\nu}=\boldsymbol{0}$, 则此分布是中心 χ^2 分布.*

证明 设 $\boldsymbol{C}$ 是满足 $\boldsymbol{CTC}'=\boldsymbol{I}$ 的非奇异矩阵, 定义 $\boldsymbol{Z}=\boldsymbol{CY}$. 则 $\boldsymbol{Z}$ 服从正态分布, 其均值为 $\mathrm{E}(\boldsymbol{Z})=\boldsymbol{C}\mathrm{E}(\boldsymbol{Y})=\boldsymbol{C\nu}=\boldsymbol{\lambda}$, 协方差阵为 $\mathrm{E}[(\boldsymbol{Z}-\boldsymbol{\lambda})(\boldsymbol{Z}-\boldsymbol{\lambda})']=\mathrm{E}[\boldsymbol{C}(\boldsymbol{Y}-\boldsymbol{\nu})(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{C}']=\boldsymbol{CTC}'=\boldsymbol{I}$. 则 $\boldsymbol{Y}'\boldsymbol{T}^{-1}\boldsymbol{Y}=\boldsymbol{Z}'(\boldsymbol{C}')^{-1}\boldsymbol{T}^{-1}\boldsymbol{C}^{-1}\boldsymbol{Z}=\boldsymbol{Z}'(\boldsymbol{CTC}')^{-1}\boldsymbol{Z}=\boldsymbol{Z}'\boldsymbol{Z}$, 即 $\boldsymbol{Z}$ 的分量的平方和. 类似地, $\boldsymbol{\nu}'\boldsymbol{T}^{-1}\boldsymbol{\nu}=\boldsymbol{\lambda}'\boldsymbol{\lambda}$. 因此, $\boldsymbol{Y}'\boldsymbol{T}^{-1}\boldsymbol{Y}$ 与 $\sum_{i=1}^m \boldsymbol{Z}_i^2$ 同分布, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_m$ 是独立的, 分别服从均值为 $\boldsymbol{\lambda}_1,\cdots,\boldsymbol{\lambda}_m$, 方差为 1 的正态分布. 由定义知, 这个分布是非中心参数为 $\sum_{i=1}^m\lambda_i^2$ 的非中心 χ^2 分布. 见 3.3.3 节. 如果 $\boldsymbol{\lambda}_1=\cdots=\boldsymbol{\lambda}_m=0$, 则这个分布是中心的. (见习题 7.5.) ■

因为 $\sqrt{N}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})$ 服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$, 所以由定理可得

$$N(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}) \tag{12}$$

服从自由度为 p 的 (中心) χ^2 分布. 这是我们设置关于 $\boldsymbol{\mu}$ 的检验和置信区域的基础事实.

设 $\chi_p^2(\alpha)$ 是使得

$$\Pr\{\chi_p^2>\chi_p^2(\alpha)\}=\alpha \tag{13}$$

的数值.

则有

$$\Pr\left\{N(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu})>\chi_p^2(\alpha)\right\}=\alpha. \tag{14}$$

为了检验假设 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$, 其中 $\boldsymbol{\mu}_0$ 是已知向量, 我们用

$$N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)>\chi_p^2(\alpha) \tag{15}$$

作为临界区域. 如果我们得到的样本满足 (15), 那么拒绝原假设. 可以直观看出, 如果 $\boldsymbol{\mu}$ 与 $\boldsymbol{\mu}_0$ 的差别非常大, 那么拒绝假设的概率大于 α, 这是因为在 $\bar{\boldsymbol{x}}$ 的空间里, (15) 定义了一个中心在 $\boldsymbol{\mu}_0$ 的椭球, 并且当 $\boldsymbol{\mu}$ 远离 $\boldsymbol{\mu}_0$ 时, $\bar{\boldsymbol{x}}$ 的密度就集中在椭球边界上的一点或界外的一点. 若 $\bar{\boldsymbol{X}}$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$、容量为 N 的样本的均值, 则 $N(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)$ 服从自由度为 p、非中心参数为 $N(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$ 的非中心 χ^2 分布. [Bose (1936a), (1936b) 给出.] Pearson (1900) 第一次证明了定理 3.3.3 在 $\boldsymbol{\nu}=\mathbf{0}$ 时的情形.

现在基于均值为 $\bar{\boldsymbol{x}}$ 的样本, 考虑下面的陈述: “分布的均值满足关于 $\boldsymbol{\mu}^*$ 的不等式

$$N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)\leqslant\chi_p^2(\alpha).\text{”} \tag{16}$$

我们从 (14) 中看出, 所抽取的样本使得上面陈述成立的概率是 $1-\alpha$, 因为 (14) 里的事件等价于上面陈述不成立. 因此, 满足 (16) 的 $\boldsymbol{\mu}^*$ 的集合是 $\boldsymbol{\mu}$ 的一个置信度为 $1-\alpha$ 的置信区域.

在 $\bar{\boldsymbol{x}}$ 的 p 维空间里, (15) 是一个中心在 $\boldsymbol{\mu}_0$ 的椭球的表面及外部, 椭球的形状依赖于 $\boldsymbol{\Sigma}^{-1}$, 给定 $\boldsymbol{\Sigma}^{-1}$ 时其大小依赖于 $(1/N)\chi_p^2(\alpha)$ 的值. 在 $\boldsymbol{\mu}^*$ 的 p 维空间里, (16) 是一个中心在 $\bar{\boldsymbol{x}}$ 的椭球的表面及内部. 如果 $\boldsymbol{\Sigma}^{-1}=\boldsymbol{I}$, 则 (14) 就表示 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{\mu}$ 的距离大于 $\sqrt{\chi_p^2(\alpha)/N}$ 的概率是 α.

定理 3.3.4 如果 $\bar{\boldsymbol{x}}$ 是取自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$、容量为 N 的样本的均值, 并且 $\boldsymbol{\Sigma}$ 已知, 则 (15) 给出了假设检验 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ 的一个大小为 α 的临界区域, (16) 给出了 $\boldsymbol{\mu}$ 的一个置信度为 $1-\alpha$ 的置信区域. 这里 $\chi_p^2(\alpha)$ 满足 (13).

这个方法还可以应用在相应的二样本问题上. 假设样本 $\{\boldsymbol{x}_\alpha^{(1)}\},\alpha=1,\cdots,N_1$, 取自分布 $N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma})$, 样本 $\{\boldsymbol{x}_\alpha^{(2)}\},\alpha=1,\cdots,N_2$, 取自分布 $N(\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma})$, 这两个总体分布有相同的协方差阵. 则这两个样本均值

$$\bar{\boldsymbol{x}}^{(1)}=\frac{1}{N_1}\sum_{\alpha=1}^{N_1}\boldsymbol{x}_\alpha^{(1)},\qquad \bar{\boldsymbol{x}}^{(2)}=\frac{1}{N_2}\sum_{\alpha=1}^{N_2}\boldsymbol{x}_\alpha^{(2)} \tag{17}$$

独立, 它们分别服从分布 $N[\boldsymbol{\mu}^{(1)},(1/N_1)\boldsymbol{\Sigma}]$ 和 $N[\boldsymbol{\mu}^{(2)},(1/N_2)\boldsymbol{\Sigma}]$. 两个样本均值的差 $\boldsymbol{y}=\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)}$ 服从分布 $N\{\boldsymbol{\nu},[(1/N_1)+(1/N_2)]\boldsymbol{\Sigma}\}$, 其中 $\boldsymbol{\nu}=\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}$. 因此

$$\frac{N_1N_2}{N_1+N_2}(\boldsymbol{y}-\boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\nu})\leqslant\chi_p^2(\alpha) \tag{18}$$

是两均值向量之差 $\boldsymbol{\nu}$ 的一个置信区域, 从而假设检验 $\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}$ 的临界区域是

$$\frac{N_1N_2}{N_1+N_2}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})>\chi_p^2(\alpha). \tag{19}$$

Mahalanobis(1930) 提出用 $(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$ 作为两个总体距离平方的一个测度. 设矩阵 $\boldsymbol{C}$ 满足 $\boldsymbol{\Sigma}=\boldsymbol{CC}'$ 且设 $\boldsymbol{\nu}^{(i)}=\boldsymbol{C}^{-1}\boldsymbol{\mu}^{(i)}, i=1,2$. 则距离平方就是欧氏距离平方 $(\boldsymbol{\nu}^{(1)}-\boldsymbol{\nu}^{(2)})'(\boldsymbol{\nu}^{(1)}-\boldsymbol{\nu}^{(2)})$.

3.3.3 非中心 χ^2 分布, 功效函数

可以用非中心 χ^2 分布来评估原假设为 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ 的检验 (15) 的功效函数. 中心 χ^2 分布是均值为 0、方差为 1 的独立 (标量) 正态变量平方和的分布, 非中心 χ^2 分布是均值不再为 0 时的推广. 设 $\boldsymbol{Y}$ (p 维) 服从分布 $N(\boldsymbol{\lambda},\boldsymbol{I})$. 设 $\boldsymbol{Q}$ 是一个正交矩阵, 其第一行元素是

$$q_{1i}=\frac{\lambda_i}{\sqrt{\boldsymbol{\lambda}'\boldsymbol{\lambda}}},\quad i=1,\cdots,p. \tag{20}$$

则 $\boldsymbol{Z}=\boldsymbol{QY}$ 服从分布 $N(\boldsymbol{\tau},\boldsymbol{I})$, 其中

$$\boldsymbol{\tau}=\begin{pmatrix}\tau\\0\\\vdots\\0\end{pmatrix}, \tag{21}$$

$\tau=\sqrt{\boldsymbol{\lambda}'\boldsymbol{\lambda}}$. 设 $V=\boldsymbol{Y}'\boldsymbol{Y}=\boldsymbol{Z}'\boldsymbol{Z}=\sum_{i=1}^p Z_i^2$. 则 $W=\sum_{i=2}^p Z_i^2$ 服从自由度为 $p-1$ 的 χ^2 分布 (习题 7.5), 并且 Z_1 和 W 有联合密度

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(z_1-\tau)^2}\frac{1}{2^{\frac{1}{2}(p-1)}\Gamma[\frac{1}{2}(p-1)]}w^{\frac{1}{2}(p-1)-1}\mathrm{e}^{-\frac{1}{2}w}\\&=C\mathrm{e}^{-\frac{1}{2}(\tau^2+z_1^2+w)}w^{\frac{1}{2}(p-3)}\mathrm{e}^{\tau z_1}\\&=C\mathrm{e}^{-\frac{1}{2}(\tau^2+z_1^2+w)}w^{\frac{1}{2}(p-3)}\sum_{\alpha=0}^{\infty}\frac{\tau^\alpha z_1^\alpha}{\alpha!},\end{aligned} \tag{22}$$

其中 $C^{-1}=2^{\frac{1}{2}p}\sqrt{\pi}\Gamma[\frac{1}{2}(p-1)]$. 做替换 $w=v-z_1^2$ (雅可比行列式是 1), $V=W+Z_1^2$ 和 Z_1 的联合密度是

$$C\mathrm{e}^{-\frac{1}{2}(\tau^2+v)}(v-z_1^2)^{\frac{1}{2}(p-3)}\sum_{\alpha=0}^{\infty}\frac{\tau^\alpha z_1^\alpha}{\alpha!}. \tag{23}$$

V 和 $U=Z_1/\sqrt{V}$ 的联合密度是 $(\mathrm{d}z_1=\sqrt{v}\mathrm{d}u)$

$$C\mathrm{e}^{-\frac{1}{2}(\tau^2+v)}v^{\frac{1}{2}(p-2)}(1-u^2)^{\frac{1}{2}(p-3)}\sum_{\alpha=0}^{\infty}\frac{\tau^\alpha v^{\frac{1}{2}\alpha}u^\alpha}{\alpha!}. \tag{24}$$

给定 v 时 z_1 的容许范围是 $-\sqrt{v}$ 到 $\sqrt{v}$, u 的容许范围是 -1 到 1. 当我们对 (24) 关于 u 逐项积分时, α 为奇数的项的积分为 0, 因为这样的项是 u 的奇函数. 在其他积分中, 我们作替换 $u=\sqrt{s}$ $(\mathrm{d}u=\frac{1}{2}\mathrm{d}s/\sqrt{s})$, 并由贝塔函数和伽玛函数的性质得到

$$\int_{-1}^{1}(1-u^2)^{\frac{1}{2}(p-3)}u^{2\beta}\mathrm{d}u = 2\int_{0}^{1}(1-u^2)^{\frac{1}{2}(p-3)}u^{2\beta}\mathrm{d}u \tag{25}$$
$$= \int_0^1 (1-s)^{\frac{1}{2}(p-3)}s^{\beta-\frac{1}{2}}\mathrm{d}s$$
$$= B\left[\frac{1}{2}(p-1), \beta+\frac{1}{2}\right]$$
$$= \frac{\Gamma[\frac{1}{2}(p-1)]\Gamma(\beta+\frac{1}{2})}{\Gamma(\frac{1}{2}p+\beta)}.$$

因此 V 的密度是

$$\frac{1}{2^{\frac{1}{2}p}\sqrt{\pi}}\mathrm{e}^{-\frac{1}{2}(\tau^2+v)}v^{\frac{1}{2}p-1}\sum_{\beta=0}^{\infty}\frac{(\tau^2)^{\beta}v^{\beta}}{(2\beta)!}\frac{\Gamma(\beta+\frac{1}{2})}{\Gamma(\frac{1}{2}p+\beta)}. \tag{26}$$

我们利用伽玛函数 (Γ 函数) 的倍量公式 $\Gamma(2\beta+1)=(2\beta)!$ (习题 7.37),

$$\Gamma(2\beta+1) = \Gamma\left(\beta+\frac{1}{2}\right)\Gamma(\beta+1)2^{2\beta}/\sqrt{\pi}, \tag{27}$$

将 (26) 重写为

$$\frac{1}{2^{\frac{1}{2}p}}\mathrm{e}^{-\frac{1}{2}(\tau^2+v)}v^{\frac{1}{2}p-1}\sum_{\beta=0}^{\infty}\left(\frac{\tau^2}{4}\right)^{\beta}\frac{1}{\beta!\Gamma(\frac{1}{2}p+\beta)}v^{\beta}. \tag{28}$$

这是自由度为 p、非中心参数为 τ^2 的非中心 χ^2 分布的密度.

定理 3.3.5 *如果 p 维向量 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{\lambda},\boldsymbol{I})$, 则 $V=\boldsymbol{Y}'\boldsymbol{Y}$ 有密度 (28), 其中 $\tau^2=\boldsymbol{\lambda}'\boldsymbol{\lambda}$.*

为了得到检验 (15) 的功效函数, 只要注意到 $\sqrt{N}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)$ 服从分布 $N[\sqrt{N}(\boldsymbol{\mu}-\boldsymbol{\mu}_0),\boldsymbol{\Sigma}]$ 即可. 由定理 3.3.3, 我们得到下面的推论.

推论 3.3.1 *如果 $\bar{\boldsymbol{X}}$ 是取自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$、容量为 N 的随机样本的均值, 则 $N(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0)$ 服从自由度为 p、非中心参数为 $N(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$ 的非中心 χ^2 分布.*

3.4 均值向量的估计的理论性质

3.4.1 极大似然估计的性质

3.3.1 节说明了 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 分别是 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的无偏估计. 在这一小节, 我们将说明 $\bar{\boldsymbol{x}}$ 和 S 是充分统计量和完全统计量.

充分性

统计量 $\boldsymbol{T}$ 对于 $\boldsymbol{X}$ 的一族分布或一个参数 $\boldsymbol{\theta}$ 是*充分的*, 如果给定 $\boldsymbol{T}=\boldsymbol{t}$ 时 $\boldsymbol{X}$ 的分布不依赖 $\boldsymbol{\theta}$ [例如, Cramér (1946), 32.4 节]. 从这种意义上讲, 统计量 $\boldsymbol{T}$ 包含与全部样本 $\boldsymbol{X}$ 一样多的关于 $\boldsymbol{\theta}$ 的信息. (当然, 这个思想严格依赖所假定的分布族.)

因子分解定理. 统计量 $t(y)$ 对 θ 是充分的当且仅当密度 $f(y|\theta)$ 可以因子分解为

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = g[\boldsymbol{t}(\boldsymbol{y}),\boldsymbol{\theta}]h(\boldsymbol{y}), \tag{1}$$

其中 $g[\boldsymbol{t}(\boldsymbol{y}),\boldsymbol{\theta}]$ 和 $h(\boldsymbol{y})$ 非负, 且 $h(\boldsymbol{y})$ 不依赖 $\boldsymbol{\theta}$.

定理 3.4.1　如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的一组观测, 则 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 对于 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 是充分的. 如果 $\boldsymbol{\mu}$ 已知, $\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'$ 对于 $\boldsymbol{\Sigma}$ 是充分的. 如果 $\boldsymbol{\Sigma}$ 已知, $\bar{\boldsymbol{x}}$ 对于 $\boldsymbol{\mu}$ 是充分的.

证明　$\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 的密度是

$$\begin{aligned}&\prod_{\alpha=1}^{N} n(\boldsymbol{x}_\alpha|\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &=(2\pi)^{-\frac{1}{2}Np}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}\exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\right] \\ &=(2\pi)^{-\frac{1}{2}Np}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}\exp\left\{-\frac{1}{2}[N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})+(N-1)\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}]\right\}.\end{aligned} \tag{2}$$

(2) 的右端项关于 $\bar{\boldsymbol{x}},\boldsymbol{S},\boldsymbol{\mu},\boldsymbol{\Sigma}$ 有形式 (1), 中间项关于 $\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'$ 和 $\boldsymbol{\Sigma}$ 有形式 (1), 其中每种情况下都取 $h(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)=1$. 右项关于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{\mu}$ 也有形式 (1), 其中取 $h(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)=\exp\{-\frac{1}{2}(N-1)\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}$. ■

注意, 如果 $\boldsymbol{\Sigma}$ 已知, $\bar{\boldsymbol{x}}$ 对于 $\boldsymbol{\mu}$ 是充分的, 但是如果 $\boldsymbol{\mu}$ 已知, $\boldsymbol{S}$ 对于 $\boldsymbol{\Sigma}$ 并不是充分的.

完全性

为了证明 T^2 检验 (5.5 节) 的优良性质, 我们需要证明 $(\bar{\boldsymbol{x}},\boldsymbol{S})$ 是 $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的完全充分统计量集合.

定义 3.4.1　指标为 $\boldsymbol{\theta}$ 的 $\boldsymbol{y}$ 的分布族是完全的, 如果对于每个实值函数 $g(\boldsymbol{y})$, 若对一切 $\boldsymbol{\theta}$,

$$\mathrm{E}_{\boldsymbol{\theta}}g(\boldsymbol{y})\equiv 0, \tag{3}$$

即关于每个 $\boldsymbol{\theta}$, 除了一个 $\boldsymbol{y}$ 的零测集外, 必有 $g(\boldsymbol{y})=0$.

如果一个充分统计量集合的分布族是完全的, 那么这个集合称为完全充分集合.

定理 3.4.2　当样本取自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 时, $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的充分统计量集合 $\bar{\boldsymbol{x}},\boldsymbol{S}$ 是完全的.

证明　我们可以像 3.3 节那样从 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_n$ 的角度来定义样本, 其中 $n=N-1$. 假定对任何函数 $g(\bar{\boldsymbol{x}},\boldsymbol{A})=g(\bar{\boldsymbol{x}},n\boldsymbol{S})$, 有

$$\begin{aligned}&\int\cdots\int K|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}g\left(\bar{\boldsymbol{x}},\sum_{\alpha=1}^{n}\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'\right) \\ &\cdot\exp\left\{-\frac{1}{2}\left[N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})+\sum_{\alpha=1}^{n}\boldsymbol{z}_\alpha'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_\alpha\right]\right\}\end{aligned} \tag{4}$$

$$\cdot \mathrm{d}\bar{\boldsymbol{x}} \prod_{\alpha=1}^{n} \mathrm{d}\boldsymbol{z}_\alpha \equiv 0, \qquad \forall \boldsymbol{\mu}, \boldsymbol{\Sigma},$$

其中 $K = \sqrt{N}(2\pi)^{-\frac{1}{2}pN}, \mathrm{d}\bar{\boldsymbol{x}} = \prod_{i=1}^{p} \mathrm{d}\bar{x}_i$ 和 $\mathrm{d}\boldsymbol{z}_\alpha = \prod_{i=1}^{p} \mathrm{d}z_{i\alpha}$. 如果我们设 $\boldsymbol{\Sigma}^{-1} = \boldsymbol{I} - 2\boldsymbol{\Theta}$, 其中 $\boldsymbol{\Theta} = \boldsymbol{\Theta}'$ 且 $\boldsymbol{I} - 2\boldsymbol{\Theta}$ 正定, 又设 $\boldsymbol{\mu} = (\boldsymbol{I} - 2\boldsymbol{\Theta})^{-1}\boldsymbol{t}$, 则 (4) 是

$$\begin{aligned} 0 \equiv & \int \cdots \int K|\boldsymbol{I} - 2\boldsymbol{\Theta}|^{\frac{1}{2}N} g\left(\bar{\boldsymbol{x}}, \sum_{\alpha=1}^{n} \boldsymbol{z}_\alpha \boldsymbol{z}_\alpha'\right) \\ & \cdot \exp\left\{-\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{I} - 2\boldsymbol{\Theta})\left(\sum_{\alpha=1}^{n} \boldsymbol{z}_\alpha \boldsymbol{z}_\alpha' + N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}'\right)\right.\right. \\ & \qquad \left.\left. - 2N\boldsymbol{t}'\bar{\boldsymbol{x}} + N\boldsymbol{t}'(\boldsymbol{I} - 2\boldsymbol{\Theta})^{-1}\boldsymbol{t}\right]\right\} \mathrm{d}\bar{\boldsymbol{x}} \prod_{\alpha=1}^{n} \mathrm{d}\boldsymbol{z}_\alpha \\ = & |\boldsymbol{I} - 2\boldsymbol{\Theta}|^{\frac{1}{2}N} \exp\left\{-\frac{1}{2}N\boldsymbol{t}'(\boldsymbol{I} - 2\boldsymbol{\Theta})^{-1}\boldsymbol{t}\right\} \int \cdots \int g(\bar{\boldsymbol{x}}, \boldsymbol{B} - N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}') \\ & \cdot \exp[\mathrm{tr}\boldsymbol{\Theta}\boldsymbol{B} + \boldsymbol{t}'(N\bar{\boldsymbol{x}})] n[\bar{\boldsymbol{x}}|\boldsymbol{0}, (1/N)\boldsymbol{I}] \prod_{\alpha=1}^{n} n(\boldsymbol{z}_\alpha|\boldsymbol{0}, \boldsymbol{I}) \mathrm{d}\bar{\boldsymbol{x}} \prod_{\alpha=1}^{n} \mathrm{d}\boldsymbol{z}_\alpha, \end{aligned} \tag{5}$$

其中 $\boldsymbol{B} = \sum_{\alpha=1}^{n} \boldsymbol{z}_\alpha \boldsymbol{z}_\alpha' + N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}'$. 因此

$$\begin{aligned} 0 \equiv & \mathrm{E}\{g(\bar{\boldsymbol{x}}, \boldsymbol{B} - N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}')\exp[\mathrm{tr}\boldsymbol{\Theta}\boldsymbol{B} + \boldsymbol{t}'(N\bar{\boldsymbol{x}})]\} \\ = & \int \cdots \int g(\bar{\boldsymbol{x}}, \boldsymbol{B} - N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}')\exp[\mathrm{tr}\boldsymbol{\Theta}\boldsymbol{B} + \boldsymbol{t}'(N\bar{\boldsymbol{x}})] h(\bar{\boldsymbol{x}}, \boldsymbol{B}) \mathrm{d}\bar{\boldsymbol{x}} \mathrm{d}\boldsymbol{B}, \end{aligned} \tag{6}$$

其中 $h(\bar{\boldsymbol{x}}, \boldsymbol{B})$ 是 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{B}$ 的联合密度且 $\mathrm{d}\boldsymbol{B} = \prod_{i \leqslant j} \mathrm{d}b_{ij}$. (6) 的右端项是 $g(\bar{\boldsymbol{x}}, \boldsymbol{B} - N\bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}')h(\bar{\boldsymbol{x}}, \boldsymbol{B})$ 的拉普拉斯变换. 因为它为 0, 所以除了一个零测集外, $g(\bar{\boldsymbol{x}}, \boldsymbol{A}) = 0$. ■

有效性

如果 q 维随机向量 $\boldsymbol{Y}$ 有均值向量 $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\nu}$ 和协方差阵 $\mathrm{E}[(\boldsymbol{Y} - \boldsymbol{\nu})(\boldsymbol{Y} - \boldsymbol{\nu})'] = \boldsymbol{\Psi}$, 则

$$(\boldsymbol{y} - \boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{y} - \boldsymbol{\nu}) = q + 2 \tag{7}$$

称为 $\boldsymbol{Y}$ 的同心椭球. [见 Cramér (1946), p. 300.] 由这个椭球内部的均匀分布定义的密度和 $\boldsymbol{Y}$ 有相同的均值向量和协方差阵. (见习题 2.14.) 设 $\boldsymbol{\theta}$ 是一个 q 维参数向量, $\boldsymbol{t}$ 是基于从协方差阵为 $\boldsymbol{\Psi}$ 的分布中抽取的 N 个观测的 $\boldsymbol{\theta}$ 的无偏估计 (即 $\mathrm{E}(\boldsymbol{t}) = \boldsymbol{\theta}$). 则椭球

$$N(\boldsymbol{t} - \boldsymbol{\theta})'\mathrm{E}\left(\frac{\partial \ln f}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln f}{\partial \boldsymbol{\theta}}\right)'(\boldsymbol{t} - \boldsymbol{\theta}) = q + 2 \tag{8}$$

整个包含在 $\boldsymbol{t}$ 的同心椭球里; $\partial \ln f/\partial \boldsymbol{\theta}$ 表示该分布 (或概率函数) 的密度关于 $\boldsymbol{\theta}$ 的分量的导数的列向量. Cramér (1946, p. 495) 的讨论是从标量观测出发的, 但是它显然也适用于向量观测. 如果 (8) 是 $\boldsymbol{t}$ 的同心椭球, 则称 $\boldsymbol{t}$ 是有效的. 总体来说, (8) 的体积与同心椭球的体积比就是 $\boldsymbol{t}$ 的效率. 在多元正态分布情形下, 如果 $\boldsymbol{\theta} = \boldsymbol{\mu}$, 则

$\bar{\boldsymbol{x}}$ 是有效的. 如果 $\boldsymbol{\theta}$ 包含 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$, 则 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 的效率是 $[(N-1)/N]^{p(p+1)/2}$. 在适当的多元正态分布满足的正则条件下,

$$\mathrm{E}\left(\frac{\partial \ln f}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln f}{\partial \boldsymbol{\theta}}\right)' = -\mathrm{E}\left(\frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right). \tag{9}$$

这是一个观测的信息矩阵. Cramér-Rao 下界使得对任意无偏估计 $\boldsymbol{t}$,

$$N\mathrm{E}[(\boldsymbol{t}-\boldsymbol{\theta})(\boldsymbol{t}-\boldsymbol{\theta})'] - \left[-\mathrm{E}\left(\frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)\right]^{-1} \tag{10}$$

都是半正定的. (还有一些其他的下界.)

一致性

定义 3.4.2 向量序列 $\boldsymbol{t}_n = (t_{1n}, \cdots, t_{mn})', n = 1, 2, \cdots$, 是 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)'$ 的一致估计, 如果 $\mathrm{plim}_{n\to\infty} t_{in} = \theta_i, i = 1, \cdots, m$.

由大数定律, 如果观测向量独立同分布, 均值为 $\boldsymbol{\mu}$, 则样本均值向量 $\bar{\boldsymbol{x}}$ 的每个分量都是期望值向量 $\boldsymbol{\mu}$ 的每个分量的一致估计, 因此 $\bar{\boldsymbol{x}}$ 是 $\boldsymbol{\mu}$ 的一致估计. 这里没有正态性的要求.

由引理 3.2.1, 取 $\boldsymbol{b} = \boldsymbol{\mu}$, 则样本协方差阵的元素是

$$s_{ij} = \frac{1}{N-1}\sum_{\alpha=1}^{N}(x_{i\alpha}-\mu_i)(x_{j\alpha}-\mu_j) - \frac{N}{N-1}(\bar{x}_i-\mu_i)(\bar{x}_j-\mu_j). \tag{11}$$

第二项的概率极限为 0. 如果 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$ 独立同分布, 均值为 $\boldsymbol{\mu}$, 协方差阵为 $\boldsymbol{\Sigma}$, 则第一项的概率极限是 σ_{ij}. 从而 $\boldsymbol{S}$ 是 $\boldsymbol{\Sigma}$ 的一致估计.

渐近正态性

首先我们来证明一个多元中心极限定理.

定理 3.4.3 设 m 维向量 $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots$ 独立同分布, 均值为 $\mathrm{E}(\boldsymbol{Y})_\alpha = \boldsymbol{\nu}$, 协方差阵为 $\mathrm{E}[(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})'] = \boldsymbol{T}$. 则当 $n\to\infty$ 时, $(1/\sqrt{n})\sum_{\alpha=1}^{n}(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})$ 的极限分布是 $N(\boldsymbol{0}, \boldsymbol{T})$.

证明 设

$$\phi_n(\boldsymbol{t}, u) = \mathrm{E}\left\{\exp\left[\mathrm{i}u\boldsymbol{t}'\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})\right]\right\}, \tag{12}$$

其中 u 是标量, $\boldsymbol{t}$ 是 m 维向量. 对于固定的 $\boldsymbol{t}$, $\phi_n(\boldsymbol{t}, u)$ 可以看作是 $(1/\sqrt{n})\sum_{\alpha=1}^{n}(\boldsymbol{t}'\boldsymbol{Y}_\alpha - \mathrm{E}(\boldsymbol{t}'\boldsymbol{Y}_\alpha))$ 的特征函数. 由单变量中心极限定理 [Cramér (1946), p. 215] 知, 其极限分布是 $N(0, \boldsymbol{t}'\boldsymbol{T}\boldsymbol{t})$. 因此 (定理 2.6.4), 对每个 u 和 $\boldsymbol{t}$, 有

$$\lim_{n\to\infty}\phi_n(\boldsymbol{t}, u) = \mathrm{e}^{-\frac{1}{2}u^2\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}. \tag{13}$$

(对 $\boldsymbol{t} = \boldsymbol{0}$ 的情况有一个特别且显然的证明.) 设 $u = 1$, 则对每一个 $\boldsymbol{t}$, 有

$$\lim_{n\to\infty}\mathrm{E}\left\{\exp\left[\mathrm{i}\boldsymbol{t}'\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})\right]\right\} = \mathrm{e}^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}. \tag{14}$$

因为 $e^{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{T}\boldsymbol{t}}$ 在 $\boldsymbol{t}=\boldsymbol{0}$ 处连续, 故在 $\boldsymbol{t}=\boldsymbol{0}$ 的某个邻域, 其收敛是一致的. 定理得证. ■

现在我们来证明, 当样本容量递增时, 样本协方差阵是渐近正态分布的.

定理 3.4.4 设 $\boldsymbol{A}(n)=\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}}_N)(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}}_N)'$, 其中 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots$ 独立且都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. 取 $n=N-1$, 则 $\boldsymbol{B}(n)=(1/\sqrt{n})[\boldsymbol{A}(n)-n\boldsymbol{\Sigma}]$ 的极限分布是均值为 $\boldsymbol{0}$、协方差为

$$\mathrm{E}(b_{ij}(n)b_{kl}(n))=\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk} \tag{15}$$

的正态分布.

证明 如以前的证明, $\boldsymbol{A}(n)$ 和 $\boldsymbol{A}(n)=\sum_{\alpha=1}^{N}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 同分布, 其中 $\boldsymbol{Z}_1,\boldsymbol{Z}_2,\cdots$ 独立且都服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. 我们将 $\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 的元素排列在一个向量里, 即

$$\boldsymbol{Y}_\alpha=\begin{pmatrix} Z_{1\alpha}^2\\ Z_{1\alpha}Z_{2\alpha}\\ \vdots\\ Z_{2\alpha}^2\\ \vdots\\ Z_{p\alpha}^2\end{pmatrix}. \tag{16}$$

$\boldsymbol{Y}_\alpha$ 的矩可以从 2.6 节给出的 $\boldsymbol{Z}_\alpha$ 的矩推演出来. 我们有 $\mathrm{E}(Z_{i\alpha}Z_{j\alpha})=\sigma_{ij}$, $\mathrm{E}(Z_{i\alpha}Z_{j\alpha}Z_{k\alpha}Z_{l\alpha})=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$, $\mathrm{E}[(Z_{i\alpha}Z_{j\alpha}-\sigma_{ij})(Z_{k\alpha}Z_{l\alpha}-\sigma_{kl})]=\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$. 因此, 由 (16) 定义的向量 $\boldsymbol{Y}_\alpha$ 满足定理 3.4.3 的条件, 其中 $\boldsymbol{\nu}$ 的元素是 $\boldsymbol{\Sigma}$ 的元素按照 (16) 相似地排列在一个向量里, $\boldsymbol{T}$ 的元素如上所给. 如果 $\boldsymbol{A}(n)$ 的元素按照 (16) 相似地排列在一个向量里, 比方说向量 $\boldsymbol{W}(n)$, 则 $\boldsymbol{W}(n)-n\boldsymbol{\nu}=\sum_{\alpha=1}^{n}(\boldsymbol{Y}_\alpha-\boldsymbol{\nu})$. 由定理 3.4.3 知, $(1/\sqrt{n})[\boldsymbol{W}(n)-n\boldsymbol{\nu}]$ 有均值为 $\boldsymbol{0}$、协方差阵与 $\boldsymbol{Y}_\alpha$ 的协方差阵相同的极限正态分布. ■

如果 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots$ 独立同分布且有有限四阶矩, 则 $\boldsymbol{B}(n)$ 的元素有均值为 $\boldsymbol{0}$ 的极限正态分布, 但是 $\boldsymbol{B}(n)$ 的协方差结构将依赖于四阶矩.

3.4.2 决策理论

用决策理论来研究估计是有启发意义的. 我们回顾一些概念. x 是随机变量 X (可能是个向量) 上的一个观测, 其中 X 的分布 $\boldsymbol{P}_\theta$ 依赖于参数 θ, θ 是集合 $\boldsymbol{\Theta}$ 中的元素, 统计学家将在集合 D 上做决策 d. 一个决策函数是定义域为 X 值的集合、值域为 D 的一个函数 $\delta(x)$. 当分布为 $\boldsymbol{P}_\theta$ 时, 决策 d 的损失是一个非负函数 $L(\theta,d)$. 决策函数 (简称决策) $\delta(x)$ 的一个评估基于其风险函数

$$R(\theta,\delta)=\mathrm{E}_\theta L[\theta,\delta(X)]. \tag{17}$$

例如, 如果 d 和 θ 是单变量的, 则其损失可能是平方误差, $L(\theta,d)=(\theta-d)^2$, 则风险函数是均方误差 $\mathrm{E}_\theta[\delta(X)-\theta]^2$.

一个决策函数 $\delta(x)$ 和一个决策函数 $\delta^*(x)$一样优, 如果

$$R(\theta,\delta) \leqslant R(\theta,\delta^*), \quad \forall\theta. \tag{18}$$

如果 (18) 中不等号对至少一个 θ 值严格成立, 则称 $\delta(x)$*优于*$\delta^*(x)$. 一个决策 $\delta^*(x)$ *是不容许的*, 如果存在另一个决策 $\delta(x)$ 优于 $\delta^*(x)$. 一个决策是*容许的*, 如果关于给定的损失函数它不是不容许的 (即如果没有优于它的决策). 一类决策称为*完全的*, 如果对任意不在这个类的决策, 总有一个该类的决策优于它. 如果该类不含完全的真子类, 则称为*最小完全的*. 如果这样的最小完全类存在, 那么称它为容许决策类. 当有这样一个类时, 就不需要 (数学上) 用这个最小完全类之外的任何决策函数了. 有时候, 为了方便, 称一类决策为*本质完全类*, 如果对每一个不在该类的决策, 都存在一个该类里的决策和它一样优.

对于给定的决策, 风险函数是参数的一个函数. 如果参数被一个先验分布决定, 比如密度 $\rho(\theta)$, 则使用决策函数 $\delta(x)$ 带来的平均损失是

$$r(\rho,\delta) = \mathrm{E}_\rho R(\theta,\delta) = \mathrm{E}_p \mathrm{E}_\theta L[\theta,\delta(X)]. \tag{19}$$

给定先验密度 ρ, 可以极小化 $r(\rho,\delta)$ 的决策函数 $\delta(x)$ 为*贝叶斯决策*, $r(\rho,\delta)$ 的极小值为*贝叶斯风险*. 在一般条件下, 贝叶斯决策是容许的, 并且容许决策是贝叶斯决策或贝叶斯决策的极限. 如果给定 θ 时 X 的密度是 $f(x|\theta)$, 则 X 和 θ 的联合密度是 $f(x|\theta)\rho(\theta)$, 并且决策 $\delta(x)$ 的平均风险是

$$\begin{aligned} r(\rho,\delta) &= \int_\Theta \int_X L[\theta,\delta(x)] f(x|\theta)\rho(\theta)\mathrm{d}x\mathrm{d}\theta \\ &= \int_X \left\{ \int_\Theta L[\theta,\delta(x)] g(\theta|x)\mathrm{d}\theta \right\} f(x)\mathrm{d}x, \end{aligned} \tag{20}$$

这里

$$f(x) = \int_\Theta f(x|\theta)\rho(\theta)\mathrm{d}\theta, \qquad g(\theta|x) = \frac{f(x|\theta)\rho(\theta)}{f(x)} \tag{21}$$

分别是 X 的边缘密度和给定 x 时 θ 的后验密度. 极小化 $r(\rho,\delta)$ 的决策对于每一个 x 极小化了 (20) 右端大括号里的表达式, 即 $L[\theta,\delta(x)]$ 关于后验分布的期望. 如果 θ 和 d 是向量 ($\boldsymbol{\theta}$ 和 $\boldsymbol{d}$) 且 $L(\boldsymbol{\theta},\boldsymbol{d}) = (\boldsymbol{\theta}-\boldsymbol{d})'\boldsymbol{Q}(\boldsymbol{\theta}-\boldsymbol{d})$, 其中 $\boldsymbol{Q}$ 正定, 则

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\theta}|x} L[\boldsymbol{\theta},\boldsymbol{d}(x)] &= \mathrm{E}_{\boldsymbol{\theta}|x}[\boldsymbol{\theta}-\mathrm{E}(\boldsymbol{\theta}|x)]'\boldsymbol{Q}[\boldsymbol{\theta}-\mathrm{E}(\boldsymbol{\theta}|x)] \\ &\quad + [\mathrm{E}(\boldsymbol{\theta}|x)-\boldsymbol{d}(x)]'\boldsymbol{Q}[\mathrm{E}(\boldsymbol{\theta}|x)-\boldsymbol{d}(x)]. \end{aligned} \tag{22}$$

取 $\boldsymbol{d}(x) = \mathrm{E}(\boldsymbol{\theta}|x)$, 即后验分布的均值时, (22) 达到极小值.

定理 3.4.5 *如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 相互独立, 每一个 $\boldsymbol{x}_\alpha$ 都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 那么若 $\boldsymbol{\mu}$ 有先验分布 $N(\boldsymbol{\nu},\boldsymbol{\Phi})$, 则给定 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 时 $\boldsymbol{\nu}$ 的后验分布是正态的, 其均值为*

$$\boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\bar{\boldsymbol{x}} + \frac{1}{N}\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\nu}, \tag{23}$$

协方差阵为

$$\boldsymbol{\Phi} - \boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Phi}. \tag{24}$$

证明 因为 $\bar{\boldsymbol{x}}$ 对 $\boldsymbol{\mu}$ 是充分的, 所以我们只需要考虑 $\bar{\boldsymbol{x}}$ 与 $\boldsymbol{\mu}+\boldsymbol{\nu}$ 同分布的情况, 其中 $\boldsymbol{\nu}$ 服从分布 $N[\mathbf{0},(1/N)\boldsymbol{\Sigma}]$ 且和 $\boldsymbol{\mu}$ 独立. 则 $\boldsymbol{\mu}$ 和 $\bar{\boldsymbol{x}}$ 的联合分布是

$$N\left[\begin{pmatrix}\boldsymbol{\nu}\\ \boldsymbol{\nu}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Phi} & \boldsymbol{\Phi}\\ \boldsymbol{\Phi} & \boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\end{pmatrix}\right]. \tag{25}$$

给定 $\bar{\boldsymbol{x}}$ 时 $\boldsymbol{\mu}$ 的条件分布的均值是 (由定理 2.5.1)

$$\boldsymbol{\nu}+\boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu}), \tag{26}$$

简化得到 (23). ■

推论 3.4.1 *如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 独立, 每一个 $\boldsymbol{x}_\alpha$ 都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{\mu}$ 有先验分布 $N(\boldsymbol{\nu},\boldsymbol{\Phi})$, 且损失函数是 $(\boldsymbol{d}-\boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{d}-\boldsymbol{\mu})$, 则 $\boldsymbol{\mu}$ 的贝叶斯估计是 (23).*

$\boldsymbol{\mu}$ 的贝叶斯估计是 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{\mu}$ 的先验均值 $\boldsymbol{\nu}$ 的加权平均. 如果 $(1/N)\boldsymbol{\Sigma}$ 比 $\boldsymbol{\Phi}$ 小 (比如, 如果 N 比较大时), 则 $\boldsymbol{\nu}$ 的权重较小. 另一方面, 如果 $\boldsymbol{\Phi}$ 大, 即先验分布的信息相对较少, 则 $\bar{\boldsymbol{x}}$ 的权重较大. 事实上, $\boldsymbol{\Phi}$ 趋于 ∞ 意味着 $\boldsymbol{\Phi}^{-1}\to\mathbf{0}$, 估计接近 $\bar{\boldsymbol{x}}$.

决策函数 $\delta_0(x)$ 是极小化极大的, 如果

$$\sup_\theta R(\theta,\delta_0)=\inf_\delta\sup_\theta R(\theta,\delta). \tag{27}$$

定理 3.4.6 *如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 相互独立且都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 并取损失函数 $(\boldsymbol{d}-\boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{d}-\boldsymbol{\mu})$, 那么 $\bar{\boldsymbol{x}}$ 是一个极小化极大估计.*

证明 这来自于统计决策理论的一个定理: 如果一个决策 $\delta_0(x)$ 是广义贝叶斯的 [即如果对任意 ε, 存在适当的 ρ, 使得 $r(\rho,\delta_0)\leqslant r(\rho,\delta_p)+\varepsilon$, 其中 δ_p 是相应的贝叶斯决策], 并且如果 $R(\theta,\delta_0)$ 是常数, 则 δ_0 是极小化极大的. [例如, 见 Ferguson (1967), 2.11 节定理 3.] 我们得到

$$\begin{aligned}R(\boldsymbol{\mu},\bar{\boldsymbol{x}})&=\mathrm{E}[(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{Q}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})]\\&=\mathrm{E}[\mathrm{tr}\boldsymbol{Q}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})(\bar{\boldsymbol{x}}-\boldsymbol{\mu})']\\&=\frac{1}{N}\mathrm{tr}\boldsymbol{Q}\boldsymbol{\Sigma}.\end{aligned} \tag{28}$$

令 (23) 为 $d(\bar{\boldsymbol{x}})$. 它的平均风险是

$$\begin{aligned}&\mathrm{E}_{\bar{\boldsymbol{x}}}\mathrm{E}_{\boldsymbol{\mu}}\{\mathrm{tr}\boldsymbol{Q}[\boldsymbol{d}(\bar{\boldsymbol{x}})-\boldsymbol{\mu}][\boldsymbol{d}(\bar{\boldsymbol{x}})-\boldsymbol{\mu}]'|\bar{\boldsymbol{x}}\}\\&=\mathrm{E}_{\bar{\boldsymbol{x}}}\mathrm{tr}\boldsymbol{Q}\left[\boldsymbol{\Phi}-\boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Phi}\right]=\mathrm{tr}\boldsymbol{Q}\boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\frac{1}{N}\boldsymbol{\Sigma}\\&=\mathrm{tr}\boldsymbol{Q}\left(\boldsymbol{I}+\frac{1}{N}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\right)^{-1}\frac{1}{N}\boldsymbol{\Sigma}\to\frac{1}{N}\mathrm{tr}\boldsymbol{Q}\boldsymbol{\Sigma}\quad(\text{当 }\boldsymbol{\Phi}^{-1}\to\mathbf{0}\text{时}).\end{aligned} \tag{29}$$

■

更多关于决策理论的讨论见 Ferguson (1967), DeGroot(1970) 或 Berger (1980b).

3.5 均值的改良估计

3.5.1 引言

样本均值 $\bar{x}$ 似乎是基于取自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本的总体均值的一个自然估计. 它是极大似然估计、$\boldsymbol{\Sigma}$ 已知时的充分统计量, 以及方差极小无偏估计. 而且它是同变的, 即如果将任意一个向量 $\boldsymbol{\nu}$ 分别加在每个观测向量和 $\boldsymbol{\mu}$ 上, 估计的误差 $\overline{(\boldsymbol{x}+\boldsymbol{\nu})}-(\boldsymbol{\mu}+\boldsymbol{\nu})=(\bar{\boldsymbol{x}}-\boldsymbol{\mu})$ 与 $\boldsymbol{\nu}$ 独立; 换句话说, 误差和原点的选择无关. 但是, Stein (1956b) 揭示了一个令人吃惊的事实: 当 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 且 $p\geqslant 3$ 时, 若取损失函数为估计分量的均方误差和, 则这个传统估计是不容许的. James and Stein (1961) 找到一个均方误差和更小的估计, 这个估计将在 3.5.2 节讨论. 随后的研究证明了这个现象是普遍且必然的.

3.5.2 James-Stein 估计

损失函数

$$L(\boldsymbol{\mu},\boldsymbol{m})=(\boldsymbol{m}-\boldsymbol{\mu})'(\boldsymbol{m}-\boldsymbol{\mu})=\sum_{i=1}^{p}(m_i-\mu_i)^2=\|\boldsymbol{m}-\boldsymbol{\mu}\|^2 \tag{1}$$

是估计分量的均方误差和. 我们将通过给出另一个对每一个均值向量 $\boldsymbol{\mu}$ 都有较小期望损失的估计 [James and Stein (1961)] 来说明样本均值是不容许的. 我们假定被抽样的正态分布的协方差阵以已知常数比例等比于 $\boldsymbol{I}$. 为了方便, 将这个常数选为使得 $\boldsymbol{Y}=(1/N)\sum_{\alpha=1}^{N}\boldsymbol{X}_\alpha=\bar{\boldsymbol{X}}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{I})$ 的数. 则估计 $\boldsymbol{Y}$ 的损失或风险期望就是 $\mathrm{E}\|\boldsymbol{Y}-\boldsymbol{\mu}\|^2=\mathrm{tr}\boldsymbol{I}=p$. James 和 Stein 提出的估计量是 (本质上)

$$\boldsymbol{m}(\boldsymbol{y})=\left(1-\frac{p-2}{\|\boldsymbol{y}-\boldsymbol{\nu}\|^2}\right)(\boldsymbol{y}-\boldsymbol{\nu})+\boldsymbol{\nu}, \tag{2}$$

其中 $\boldsymbol{\nu}$ 是任意固定向量且 $p\geqslant 3$. 这个估计收缩了观测 $\boldsymbol{y}$ 与特定 $\boldsymbol{\nu}$ 的距离. 如果 $\boldsymbol{y}$ 与 $\boldsymbol{\nu}$ 非常不同, 则收缩量是可以忽略的; 如果 $\boldsymbol{y}$ 与 $\boldsymbol{\nu}$ 比较近, 则收缩量是相当大的. 从这个角度看, $\boldsymbol{\nu}$ 是一个有利点.

定理 3.5.1 当 $p\geqslant 3$ 时, 关于损失函数 (1), 估计 (2) 的风险小于估计 $\boldsymbol{Y}$ 的风险.

我们将用下面 Stein (1974) 提出的引理来证明估计 $\boldsymbol{Y}$ 的风险减去估计 (2) 的风险是正数.

引理 3.5.1 如果函数 $f(x)$ 满足对所有 a 和 $b(a<b)$ 有

$$f(b)-f(a)=\int_a^b f'(x)\mathrm{d}x, \tag{3}$$

且若

$$\int_{-\infty}^{\infty}|f'(x)|\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x<\infty, \tag{4}$$

则

$$\int_{-\infty}^{\infty} f(x)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x = \int_{-\infty}^{\infty} f'(x)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x. \tag{5}$$

引理的证明 我们将 (5) 的左端写为

$$\begin{aligned}
&\int_{\theta}^{\infty}[f(x)-f(\theta)](x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x \qquad (6)\\
&\quad+\int_{-\infty}^{\theta}[f(x)-f(\theta)](x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x\\
&=\int_{\theta}^{\infty}\int_{\theta}^{x} f'(y)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}y\mathrm{d}x\\
&\quad-\int_{-\infty}^{\theta}\int_{x}^{\theta} f'(y)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}y\mathrm{d}x\\
&=\int_{\theta}^{\infty}\int_{y}^{\infty} f'(y)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x\mathrm{d}y\\
&\quad-\int_{-\infty}^{\theta}\int_{-\infty}^{y} f'(y)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2}\mathrm{d}x\mathrm{d}y,
\end{aligned}$$

它等于 (5) 的右端. Fubini 定理解释了积分次序的交换. (见习题 3.22.) ■

在某些特别情况下, 这个引理也可以通过分部积分的方法得到.

定理 3.5.1 的证明 风险差是

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &= \mathrm{E}_{\boldsymbol{\mu}}\{\|\boldsymbol{Y}-\boldsymbol{\mu}\|^2-\|\boldsymbol{m}(\boldsymbol{Y})-\boldsymbol{\mu}\|^2\} \qquad (7)\\
&=\mathrm{E}_{\boldsymbol{\mu}}\left\{\|\boldsymbol{Y}-\boldsymbol{\mu}\|^2-\left\|\left(1-\frac{p-2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\right)(\boldsymbol{Y}-\boldsymbol{\nu})+\boldsymbol{\nu}-\boldsymbol{\mu}\right\|^2\right\}\\
&=\mathrm{E}_{\boldsymbol{\mu}}\left\{\sum_{i=1}^{p}(Y_i-\mu_i)^2-\sum_{i=1}^{p}\left[(Y_i-\mu_i)-\frac{p-2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}(Y_i-\nu_i)\right]^2\right\}\\
&=\mathrm{E}_{\boldsymbol{\mu}}\left\{2\frac{p-2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\sum_{i=1}^{p}(Y_i-\mu_i)(Y_i-\nu_i)-\frac{(p-2)^2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\right\}.
\end{aligned}$$

现在我们利用引理 3.5.1, 取

$$f(y_i)=\frac{y_i-\nu_i}{\sum_{j=1}^{p}(y_j-\nu_j)^2},\quad f'(y_i)=\frac{1}{\sum_{j=1}^{p}(y_j-\nu_j)^2}-\frac{2(y_i-\nu_i)^2}{\left[\sum_{j=1}^{p}(y_j-\nu_j)^2\right]^2}. \tag{8}$$

[对 $p\geqslant 3$, 条件 (4) 满足.] 则 (7) 为

$$\begin{aligned}
\Delta R(\boldsymbol{\mu}) &= \mathrm{E}_{\boldsymbol{\mu}}\left\{2(p-2)\sum_{i=1}^{p}\left[\frac{1}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}-\frac{2(y_i-\nu_i)^2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^4}\right]-\frac{(p-2)^2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\right\} \qquad (9)\\
&=(p-2)^2\mathrm{E}_{\boldsymbol{\mu}}\frac{1}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}>0.
\end{aligned}$$

■

这个定理说明, 当 $p\geqslant 3$ 时, $\boldsymbol{Y}$ 对 $\boldsymbol{\mu}$ 的估计是不容许的, 因为对任意 $\boldsymbol{\mu}$, 估计 (2) 有更小的风险 (与 $\boldsymbol{\nu}$ 的选取无关).

风险是均方误差和 $\mathrm{E}[m_i(\boldsymbol{Y})-\mu_i]^2$. 既然 $Y_1,\cdots,Y_p$ 独立且仅 Y_i 的分布依赖 μ_i, 那么为什么还要用所有的 Y_j 来估计 μ_i, 好像用了不相关的信息. Stein 将这个现象解释为, $\boldsymbol{Y}$ 到 $\boldsymbol{\nu}$ 的平方样本距离, 即 $\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2$, 过度估计了 $\boldsymbol{\mu}$ 到 $\boldsymbol{\nu}$ 的平方距离, 因此可以使估计 $\boldsymbol{Y}$ 更靠近 $\boldsymbol{\nu}$ 从而改进它 (无论 $\boldsymbol{\nu}$ 是什么). 在 Brown 之后, Berger (1980a) 用图 3.4 来解释. 四个点 $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, $\boldsymbol{x}_4$ 代表一个中心在 $\boldsymbol{\mu}$ 的球面分布. 考虑收缩效应. 如果收缩是某个定量, $\boldsymbol{m}(\boldsymbol{x}_1)$ 和 $\boldsymbol{m}(\boldsymbol{x}_3)$ 到 $\boldsymbol{\mu}$ 的平均距离, 比 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_3$ 到 $\boldsymbol{\mu}$ 的平均距离稍微大些, 但是 $\boldsymbol{m}(\boldsymbol{x}_2)$ 和 $\boldsymbol{m}(\boldsymbol{x}_4)$ 到 $\boldsymbol{\mu}$ 的平均距离, 比 $\boldsymbol{x}_2$ 和 $\boldsymbol{x}_4$ 到 $\boldsymbol{\mu}$ 的平均距离稍微小些. 如果 $p=3$, 另有两个点 (不在 $\boldsymbol{\nu}$ 和 $\boldsymbol{\mu}$ 的直线上) 被压缩地距 $\boldsymbol{\mu}$ 更近.

估计 (2) 的风险是

$$\mathrm{E}_{\boldsymbol{\mu}}\|\boldsymbol{m}(\boldsymbol{Y})-\boldsymbol{\mu}\|^2=p-(p-2)^2\mathrm{E}_{\boldsymbol{\mu}}\frac{1}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}, \tag{10}$$

其中 $\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2$ 服从自由度为 p、非中心参数为 $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|^2$ 的非中心 χ^2 分布. $\boldsymbol{\mu}$ 离 $\boldsymbol{\nu}$ 越远, James-Stein 估计改进得就越小, 但是总是有改进的. $\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2=V$ 的密度, 比如说是 3.3.3 节的 (28), 其中 $\tau^2=\|\boldsymbol{\mu}-\boldsymbol{\nu}\|^2$. 则对 $p\geqslant 3$,

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\mu}}\frac{1}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}&=\mathrm{E}_{\tau^2}V^{-1} \\ &=\mathrm{e}^{-\frac{1}{2}\tau^2}2^{-\frac{1}{2}p}\sum_{\beta=0}^{\infty}\left\{\frac{\tau^2}{4}\right\}^{\beta}\frac{1}{\beta!\Gamma(\frac{1}{2}p+\beta)}\int_0^{\infty}v^{\frac{1}{2}p+\beta-2}\mathrm{e}^{-\frac{1}{2}v}\mathrm{d}v \\ &=\mathrm{e}^{-\frac{1}{2}\tau^2}2^{-\frac{1}{2}p}\sum_{\beta=0}^{\infty}\left\{\frac{\tau^2}{4}\right\}^{\beta}\frac{\Gamma(\frac{1}{2}p+\beta-1)2^{\frac{1}{2}p+\beta-1}}{\beta!\Gamma(\frac{1}{2}p+\beta)} \\ &=\frac{1}{2}\mathrm{e}^{-\frac{1}{2}\tau^2}\sum_{\beta=0}^{\infty}\left\{\frac{\tau^2}{2}\right\}^{\beta}\frac{1}{\beta!(\frac{1}{2}p+\beta-1)}.\end{aligned} \tag{11}$$

注意到, 对 $\boldsymbol{\mu}=\boldsymbol{\nu}$, 即 $\tau^2=0$, (11) 是 $1/(p-2)$ 且 (10) 的均方误差是 2. 对于较大的 p, 风险减少很多.

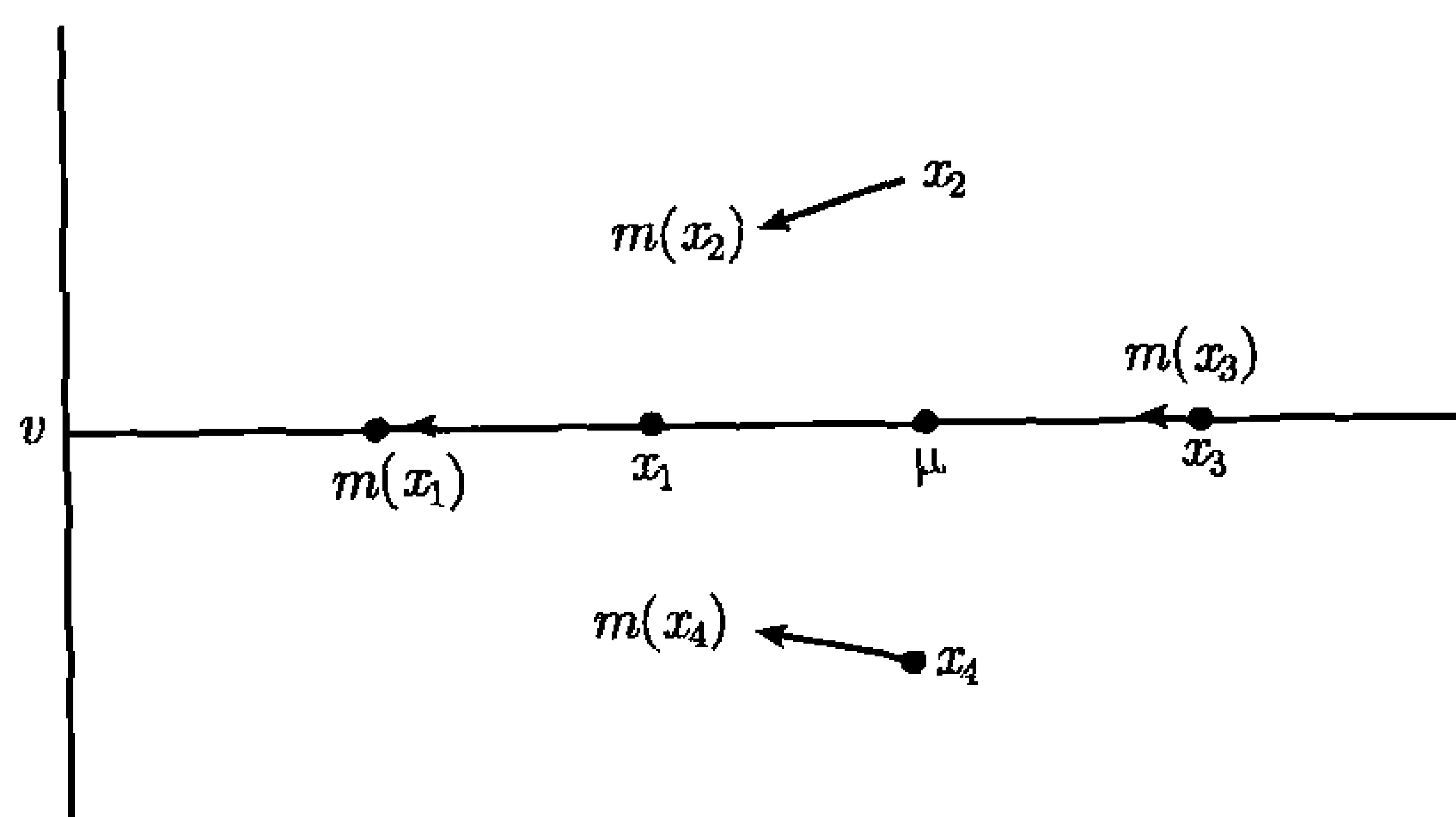


图 3.4 收缩效应

表 3.2 给出了 $p=10,\sigma^2=1$ 的风险值. 例如, 如果 $\tau^2=\|\boldsymbol{\mu}-\boldsymbol{\nu}\|^2$ 是 5, 则 James-Stein 估计的均方误差是 8.86, 相对应的自然估计的均方误差是 10. 这是例如 $\mu_i-\nu_i=1/\sqrt{2}=0.707(i=1,\cdots,10)$ 时的情形.

表 3.2[①] **James-Stein 估计的均方误差平均, $p=10,\sigma^2=1$**

$\tau^2=\|\mu-\boldsymbol{\nu}\|^2$	$\mathrm{E}_{\mu}\|m(Y)-\mu\|^2$
0.0	2.00
0.5	4.78
1.0	6.21
2.0	7.51
3.0	8.24
4.0	8.62
5.0	8.86
6.0	9.03

使用这类估计的一个明显问题是, 对于要收缩的观测均值向量, 怎么选择 $\boldsymbol{\nu}$, 任何 $\boldsymbol{\nu}$ 都能产生一个比自然估计好的估计. 但是, 就像从表 3.2 中看到的, 如果 $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|^2$ 非常大的话, 改进是很小的. 因此, 为了使改进更有效, 需要知道一些关于 $\boldsymbol{\mu}$ 的位置的信息. 这个方法的缺点就是它不是客观的, $\boldsymbol{\nu}$ 的选择依赖于研究者.

研究发现, 这个估计有一个看似缺点的特点, 即对于较小的 $\|\boldsymbol{Y}-\boldsymbol{\nu}\|$, 乘子 $\boldsymbol{Y}-\boldsymbol{\nu}$ 是负的, 也就是说, 估计 $\boldsymbol{m}(\boldsymbol{Y})$ 相对 $\boldsymbol{\nu}$ 的方向, 与 $\boldsymbol{Y}$ 相对 $\boldsymbol{\nu}$ 的相反. 这个缺点可以克服, 当因子是负数时通过用 0 代替它来改进估计.

定义 3.5.1 对任意函数 $g(u)$, 设

$$g^+(u)=\begin{cases} g(u), & g(u)\geqslant 0, \\ 0, & g(u)<0. \end{cases} \tag{12}$$

引理 3.5.2 当 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{I})$ 时,

$$\mathrm{E}_{\boldsymbol{\mu}}\{\|g^+(\|\boldsymbol{X}\|)\boldsymbol{X}-\boldsymbol{\mu}\|^2\}\leqslant \mathrm{E}_{\boldsymbol{\mu}}\{\|g(\|\boldsymbol{X}\|)\boldsymbol{X}-\boldsymbol{\mu}\|^2\}. \tag{13}$$

证明 (13) 的右端项减去左端项是,

$$\mathrm{E}_{\boldsymbol{\mu}}\{g^2(\|\boldsymbol{X}\|)\|\boldsymbol{X}\|^2-[g^+(\|\boldsymbol{X}\|)]^2\|\boldsymbol{X}\|^2\}\geqslant 0 \tag{14}$$

加上 2 倍的

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{\mu}}\boldsymbol{\mu}'\boldsymbol{X}[g^+(\|\boldsymbol{X}\|)-g(\|\boldsymbol{X}\|)] \\ &=\|\boldsymbol{\mu}\|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}y_1[g^+(\|\boldsymbol{y}\|)-g(\|\boldsymbol{y}\|)] \\ &\quad\cdot\frac{1}{(2\pi)^{\frac{1}{2}p}}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{p}y_i^2-2y_1\|\boldsymbol{\mu}\|+\|\boldsymbol{\mu}\|^2\right]\right\}\mathrm{d}\boldsymbol{y}, \end{aligned} \tag{15}$$

① 选自 Efron and Morris (1977).

其中 $\boldsymbol{y}' = \boldsymbol{x}'\boldsymbol{P}$, $(\|\boldsymbol{\mu}\|, 0, \cdots, 0) = \boldsymbol{\mu}'\boldsymbol{P}$, 且 $\boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$. [$\boldsymbol{P}$ 的第一列是 $(1/\|\boldsymbol{\mu}\|)\boldsymbol{\mu}$.] 则 (15) 是 $\|\boldsymbol{\mu}\|$ 倍的

$$\begin{aligned} &\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{\mu}\|^2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{0}^{\infty} y_1[g^+(\|\boldsymbol{y}\|) - g(\|\boldsymbol{y}\|)][\mathrm{e}^{\|\boldsymbol{\mu}\|y_1} - \mathrm{e}^{-\|\boldsymbol{\mu}\|y_1}] \\ &\cdot \frac{1}{(2\pi)^{\frac{1}{2}p}}\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{p} y_i^2}\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_p \geqslant 0 \end{aligned} \tag{16}$$

(对于 $y_1 < 0$, 用 $-y_1$ 替换 y_1). ■

定理 3.5.2　估计

$$\boldsymbol{m}^+(\boldsymbol{y}) = \left(1 - \frac{p-2}{\|\boldsymbol{y} - \boldsymbol{\nu}\|^2}\right)^+ (\boldsymbol{y} - \boldsymbol{\nu}) + \boldsymbol{\nu} \tag{17}$$

比 (2) 定义的 $\boldsymbol{m}(\boldsymbol{y})$ 有更小的风险并且是极小化极大的.

证明　在引理 3.5.2 中, 设 $g(u) = 1 - (p-2)/u^2$, $\boldsymbol{X} = \boldsymbol{Y} - \boldsymbol{\nu}$, 并将 $\boldsymbol{\mu}$ 替换为 $\boldsymbol{\mu} - \boldsymbol{\nu}$. 定理中第二个结论可由定理 3.4.6 得到. ■

这个定理说明了 $\boldsymbol{m}(\boldsymbol{Y})$ 是不容许的. 然而, 已知 $\boldsymbol{m}^+(\boldsymbol{y})$ 也是不容许的, 但是可知不可能再有较大的改进.

这个方法可以简单地推广到观测 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 且损失函数为 $L(\boldsymbol{\mu}, \boldsymbol{m}) = (\boldsymbol{m} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m} - \boldsymbol{\mu})$ 的情形. 设 $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}'$, 对某个非奇异 $\boldsymbol{C}$, $\boldsymbol{x}_\alpha = \boldsymbol{C}\boldsymbol{x}_\alpha^*$, $\alpha = 1, \cdots, N$, $\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{\mu}^*$, 并且 $L^*(\boldsymbol{m}^*, \boldsymbol{\mu}^*) = \|\boldsymbol{m}^* - \boldsymbol{\mu}^*\|^2$. 则 $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_N^*$ 是来自 $N(\boldsymbol{\mu}^*, \boldsymbol{I})$ 的观测, 问题简化为早先的问题. 于是

$$\left(1 - \frac{p-2}{N(\bar{\boldsymbol{x}} - \boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\nu})}\right)^+ (\bar{\boldsymbol{x}} - \boldsymbol{\nu}) + \boldsymbol{\nu} \tag{18}$$

是 $\boldsymbol{\mu}$ 的一个极小化极大估计.

3.5.3　已知广义协方差阵和任意二次损失函数情况下的估计

设母分布是 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 其中 $\boldsymbol{\Sigma}$ 已知, 且设损失函数是

$$L(\boldsymbol{\mu}, \boldsymbol{m}) = (\boldsymbol{m} - \boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{m} - \boldsymbol{\mu}), \tag{19}$$

其中 $\boldsymbol{Q}$ 是一个任意的正定矩阵, 它反映不同方向上相对的误差重要性. (如果损失函数是奇异的, 则可通过减少 $\boldsymbol{x}$ 的维数使得损失矩阵非奇异.) 则样本均值 $\bar{\boldsymbol{x}}$ 服从分布 $N(\boldsymbol{\mu}, (1/N)\boldsymbol{\Sigma})$ 且有风险 (期望损失)

$$\mathrm{E}[(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{Q}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})] = \mathrm{E}[\mathrm{tr}\boldsymbol{Q}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'] = \frac{1}{N}\mathrm{tr}\boldsymbol{Q}\boldsymbol{\Sigma}, \tag{20}$$

这是一个常数, 它不依赖 $\boldsymbol{\mu}$.

已经出现了一些改进 $\bar{\boldsymbol{x}}$ 的估计. 首先我们来看一个由 Berger(1975) 和 Hudson(1974) 分别独立提出的估计.

定理 3.5.3　设 $r(z), 0 \leqslant z < \infty$, 是一个非降可微函数, 满足 $0 \leqslant r(z) \leqslant 2(p-2)$. 则对于 $p \geqslant 3$,

$$m = \left(I - \frac{r(N^2(\bar{x}-\nu)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(\bar{x}-\nu))}{N(\bar{x}-\nu)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(\bar{x}-\nu)}Q^{-1}\Sigma^{-1}\right)(\bar{x}-\nu)+\nu \tag{21}$$

比 $\bar{x}$ 有较小的风险且是极小化极大的.

证明 存在矩阵 C, 使得 $C'QC = I$, $(1/N)\Sigma = C\Delta C'$, 其中 Δ 是一个对角元素为 $\delta_1 \geqslant \delta_2 \geqslant \cdots \geqslant \delta_p > 0$ 的对角矩阵 (附录中的定理 A.2.2). 设 $\bar{x} = Cy + \nu$, $\mu = C\mu^* + \nu$. 则 y 服从分布 $N(\mu^*, \Delta)$, 并且变换后的损失函数是

$$L^*(m^*, \mu^*) = (m^* - \mu^*)'(m^* - \mu^*) = \|m^* - \mu^*\|^2. \tag{22}$$

μ 的估计 (21) 变换成 $\mu^* = C^{-1}(\mu - \nu)$ 的估计,

$$m^*(y) = \left(I - \frac{r(y'\Delta^{-2}y)}{y'\Delta^{-2}y}\Delta^{-1}\right)y. \tag{23}$$

我们现在像定理 3.5.1 的证明中那样推导. y 和 m^* 的风险差是

$$\begin{aligned}\Delta R(\mu^*) &= \mathrm{E}_{\mu^*}\left\{\|Y - \mu^*\|^2 - \|m^*(Y) - \mu^*\|^2\right\} \\ &= \mathrm{E}_{\mu^*}\left\{2\frac{r(Y'\Delta^{-2}Y)}{Y'\Delta^{-2}Y}\sum_{i=1}^{p}\frac{1}{\delta_i}Y_i(Y_i - \mu_i^*) - \frac{r^2(Y'\Delta^{-2}Y)}{Y'\Delta^{-2}Y}\right\}.\end{aligned} \tag{24}$$

既然 $r(z)$ 可微, 我们用引理 3.5.1, 取 $(x - \theta) = (y_i - \mu_i^*)\delta_i$,

$$f(y_i) = \frac{r(y'\Delta^{-2}y)}{y'\Delta^{-2}y}y_i, \tag{25}$$

$$f'(y_i) = \frac{r(y'\Delta^{-2}y)}{y'\Delta^{-2}y} + \frac{2r'(y'\Delta^{-2}y)}{y'\Delta^{-2}y}\frac{y_i^2}{\delta_i^2} - \frac{2r(y'\Delta^{-2}y)}{y'\Delta^{-2}y}\frac{y_i^2}{\delta_i^2}. \tag{26}$$

则

$$\Delta R(\mu^*) = \mathrm{E}_{\mu^*}\left\{2(p-2)\frac{r(Y'\Delta^{-2}Y)}{Y'\Delta^{-2}Y} + 4r'(Y'\Delta^{-2}Y) - \frac{r^2(Y'\Delta^{-2}Y)}{Y'\Delta^{-2}Y}\right\} \geqslant 0, \tag{27}$$

这是因为 $r(y'\Delta^{-2}y) \leqslant 2(p-2)$ 和 $r'(y'\Delta^{-2}y) \geqslant 0$. ■

推论 3.5.1 对 $p \geqslant 3$,

$$\left\{I - \frac{\min[p-2, N^2(\bar{x}-\nu)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(\bar{x}-\nu)]}{N(\bar{x}-\nu)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(\bar{x}-\nu)}Q^{-1}\Sigma^{-1}\right\}(\bar{x}-\nu)+\nu \tag{28}$$

比 $\bar{x}$ 有更小的风险且是极小化极大的.

证明 函数 $r(z) = \min(p-2, z)$ 在除点 $z = p-2$ 外都是可微的. 函数 $r(z)$ 可以被一个可微函数任意逼近. (比如, 在 $z = p-2$ 处的隅角可以被任意小半径的圆弧平滑.) 我们不再给出证明细节. ■

在标准型中, 用一个标量乘以一个对角矩阵来收缩 y. 分量的方差越大, 收缩的效应越小.

Berger (1975) 对于一个更普通的密度, 即混合正态, 证明了这些结果. Berger (1976) 又证明了在正态情况下, 如果对于 $3 - \frac{1}{2}p \leqslant c < 1 + \frac{1}{2}p$,

$$r(z)=\frac{z\int_0^{\alpha}u^{\frac{1}{2}p-c+1}\mathrm{e}^{-\frac{1}{2}uz}\mathrm{d}u}{\int_0^{\alpha}u^{\frac{1}{2}p-c}\mathrm{e}^{-\frac{1}{2}uz}\mathrm{d}u},\tag{29}$$

其中 α 是 $\boldsymbol{\Sigma Q}$ 的最小特征根, 则 (21) 给出的估计 $\boldsymbol{m}$ 是极小化极大的, 当 $c<2$ 时是容许的, 当 $c<1$ 时是真贝叶斯的.

Bhatacharya(1966) 介绍了另一个极小化极大估计的方法. 设 $\boldsymbol{C}$ 满足 $\boldsymbol{C}^{-1}(1/N)\boldsymbol{\Sigma}(\boldsymbol{C}^{-1})'=\boldsymbol{I}$, $\boldsymbol{C}'\boldsymbol{QC}=\boldsymbol{Q}^*$, 其中 $\boldsymbol{Q}^*$ 是一个对角元素为 $q_1^*\geqslant q_2^*\geqslant\cdots\geqslant q_p^*>0$ 的对角矩阵. 则 $\boldsymbol{y}=\boldsymbol{C}^{-1}\bar{\boldsymbol{x}}$ 服从分布 $N(\boldsymbol{\mu}^*,\boldsymbol{I})$ 且损失函数是

$$\begin{aligned}L^*(\boldsymbol{m}^*,\boldsymbol{\mu}^*)&=\sum_{i=1}^p q_i^*(m_i^*-\mu_i^*)^2\\&=\sum_{i=1}^p\sum_{j=i}^p\alpha_j(m_i^*-\mu_i^*)^2\\&=\sum_{j=1}^p\alpha_j\sum_{i=1}^j(m_i^*-\mu_i^*)^2\\&=\sum_{j=1}^p\alpha_j\|\boldsymbol{m}^{*(j)}-\boldsymbol{\mu}^{*(j)}\|^2,\end{aligned}\tag{30}$$

其中 $\alpha_j=q_j^*-q_{j+1}^*, j=1,\cdots,p-1, \alpha_p=q_p^*, \boldsymbol{m}^{*(j)}=(m_1^*,\cdots,m_j^*)'$, $\boldsymbol{\mu}^{*(j)}=(\mu_1^*,\cdots,\mu_j^*)', j=1,\cdots,p$. 这个损失函数的分解暗示要组合向量 $\boldsymbol{\mu}^{*(j)}, j=1,\cdots,p$ 的极小化极大估计. 设 $\boldsymbol{y}^{(j)}=(y_1,\cdots,y_j)'$.

定理 3.5.4 如果 $\boldsymbol{h}^{(j)}(\boldsymbol{y}^{(j)})=[h_1^{(j)}(\boldsymbol{y}^{(j)}),\cdots,h_j^{(j)}(\boldsymbol{y}^{(j)})]'$ 是 $\boldsymbol{\mu}^{*(j)}$ 在损失函数 $\|\boldsymbol{m}^{*(j)}-\boldsymbol{\mu}^{*(j)}\|^2(j=1,\cdots,p)$ 下的一个极小化极大估计, 则

$$\frac{1}{q_i^*}\sum_{j=1}^p\alpha_j h_i^{(j)}(\boldsymbol{y}^{(j)}),\quad i=1,\cdots,p,\tag{31}$$

是 $\mu_1^*,\cdots,\mu_p^*$ 的一个极小化极大估计.

证明 首先考虑对第 i 个分量的随机估计

$$\Pr\left\{G_i(\boldsymbol{y})=h_i^{(j)}(\boldsymbol{y}^{(j)})\right\}=\frac{\alpha_j}{q_i^*},\quad j=i,\cdots,p.\tag{32}$$

则这个估计的风险是

$$\begin{aligned}\sum_{i=1}^p q_i^*\mathrm{E}_{\boldsymbol{\mu}^*}[G_i(\boldsymbol{Y})-\mu_i^*]^2&=\sum_{i=1}^p q_i^*\sum_{j=i}^p\frac{\alpha_j}{q_i^*}\mathrm{E}_{\boldsymbol{\mu}^*}[h_i^{(j)}(\boldsymbol{Y}^{(j)})-\mu_i{}^*]^2\\&=\sum_{j=1}^p\alpha_j\sum_{i=1}^j\mathrm{E}_{\boldsymbol{\mu}^*}[h_i^{(j)}(\boldsymbol{Y}^{(j)})-\mu_i{}^*]^2\\&=\sum_{j=1}^p\alpha_j\mathrm{E}_{\boldsymbol{\mu}^*}\|\boldsymbol{h}^{(j)}(\boldsymbol{Y}^{(j)})-\boldsymbol{\mu}^{*(j)}\|^2\end{aligned}\tag{33}$$

$$\leqslant \sum_{j=1}^{p} \alpha_j j = \sum_{j=1}^{p} q_j^*$$
$$= \mathrm{E}_{\boldsymbol{\mu}^*} L^*(\boldsymbol{Y}, \boldsymbol{\mu}^*)^*,$$

因此 (32) 定义的估计是极小化极大的.

因为 $G_i(\boldsymbol{Y})$ 关于 (32) 的期望值是 (31) 且损失函数是凸的, 所以估计 (31) 的风险小于随机估计的风险 (由詹生 (Jensen) 不等式). ■

3.6 椭球等高分布

3.6.1 椭球等高的观测

设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是一个密度为 $|\boldsymbol{\Lambda}|^{-\frac{1}{2}} g[(\boldsymbol{x}_\alpha - \boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\nu})]$ 的随机向量 $\boldsymbol{X}$ 的 $N(= n+1)$ 个独立观测. 样本密度是

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}N} \prod_{\alpha=1}^{N} g[(\boldsymbol{x} - \boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\nu})]. \tag{1}$$

样本均值 $\bar{\boldsymbol{x}}$ 和协方差阵 $\boldsymbol{S} = (1/n)[\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})(\boldsymbol{x}_\alpha - \boldsymbol{\mu})' - N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})']$ 分别是均值 $\boldsymbol{\mu} = \boldsymbol{\nu}$ 和协方差阵 $\boldsymbol{\Sigma} = [\mathrm{E}(R^2)/p]\boldsymbol{\Lambda}$ 的无偏估计, 其中 $R^2 = (\boldsymbol{x} - \boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\nu})$.

定理 3.6.1 来自 $|\boldsymbol{\Lambda}|^{-\frac{1}{2}} g[(\boldsymbol{x} - \boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x} - \boldsymbol{\nu})]$ 且 $\mathrm{E}(R^4) < \infty$ 的容量为 N 的样本均值与样本协方差的协方差是

$$\mathrm{E}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})(\bar{\boldsymbol{x}} - \boldsymbol{\mu})' = \frac{1}{N}\boldsymbol{\Sigma}, \tag{2}$$

$$\mathrm{E}(s_{ij} - \sigma_{ij})(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) = \boldsymbol{0}, \qquad i, j = 1, \cdots, p, \tag{3}$$

$$\begin{aligned}\mathrm{E}(s_{ij} - \sigma_{ij})(s_{kl} - \sigma_{kl}) = & \frac{\kappa}{N}(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \\ & + \frac{1}{n}(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}), \qquad i, j, k, l = 1, \cdots, p.\end{aligned} \tag{4}$$

引理 3.6.1 S 的元素的二阶矩是

$$\mathrm{E}(s_{ij}s_{kl}) = \sigma_{ij}\sigma_{kl} + \frac{1}{n}(\sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + \frac{\kappa}{N}(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) \qquad i, j, k, l = 1, \cdots, p. \tag{5}$$

引理 3.6.1 的证明 我们有

$$\begin{aligned}&\mathrm{E}\left[\sum_{\alpha,\beta=1}^{N}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j)(x_{k\beta} - \mu_k)(x_{l\beta} - \mu_l)\right] \\ &= N\mathrm{E}[(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j)(x_{k\alpha} - \mu_k)(x_{l\alpha} - \mu_l)] \\ &\quad + N(N-1)\mathrm{E}[(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j)]\mathrm{E}[(x_{k\beta} - \mu_k)(x_{l\beta} - \mu_l)] \\ &= N(1+\kappa)(\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}) + N(N-1)\sigma_{ij}\sigma_{kl},\end{aligned} \tag{6}$$

$$\begin{aligned}
&\mathrm{E}[N^2(\bar{x}_i-\mu_i)(\bar{x}_j-\mu_j)(\bar{x}_k-\mu_k)(\bar{x}_l-\mu_l)] \qquad (7)\\
&=\frac{1}{N^2}\mathrm{E}\left[\sum_{\alpha,\beta,\gamma,\delta=1}^{N}(x_{i\alpha}-\mu_i)(x_{j\beta}-\mu_j)(x_{k\gamma}-\mu_k)(x_{l\delta}-\mu_l)\right]\\
&=\frac{1}{N}(1+\kappa)(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})\\
&\quad+\frac{N-1}{N}(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})
\end{aligned}$$

$$\mathrm{E}\left[\sum_{\alpha=1}^{N}(x_{i\alpha}-\mu_i)(x_{j\alpha}-\mu_j)\frac{1}{N}\sum_{\beta,\gamma=1}^{N}(x_{k\beta}-\mu_k)(x_{l\gamma}-\mu_\gamma)\right] \tag{8}$$

$$=(1+\kappa)(\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk})+(N-1)\sigma_{ij}\sigma_{kl}.$$ ■

利用更多的矩阵代数会更方便. 定义 $\mathrm{vec}\boldsymbol{B}, \boldsymbol{B}\otimes\boldsymbol{C}$(克罗内克 (Kronecker) 积), $\boldsymbol{K}_{mn}$ (换位矩阵) 为

$$\mathrm{vec}\boldsymbol{B}=\mathrm{vec}(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n)=\begin{bmatrix}\boldsymbol{b}_1\\ \vdots\\ \boldsymbol{b}_n\end{bmatrix}, \tag{9}$$

$$\boldsymbol{B}\otimes\boldsymbol{C}=\begin{bmatrix}b_{11}\boldsymbol{C} & \cdots & b_{1n}\boldsymbol{C}\\ \vdots & & \vdots\\ b_{m1}\boldsymbol{C} & \cdots & b_{mn}\boldsymbol{C}\end{bmatrix}, \tag{10}$$

$$\boldsymbol{K}_{mn}\mathrm{vec}\boldsymbol{B}=\mathrm{vec}\boldsymbol{B}'. \tag{11}$$

例如, 见 Magnus and Neudecker (1979) 或附录 A.5 节. 我们可将 (4) 重记为

$$\begin{aligned}
\mathrm{C}(\mathrm{vec}\boldsymbol{S})&=\mathrm{E}(\mathrm{vec}\boldsymbol{S}-\mathrm{vec}\boldsymbol{\Sigma})(\mathrm{vec}\boldsymbol{S}-\mathrm{vec}\boldsymbol{\Sigma})' \qquad (12)\\
&=\frac{n\kappa+N}{nN}(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\frac{\kappa}{N}\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})'.
\end{aligned}$$

定理 3.6.2

$$\sqrt{n}\begin{bmatrix}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})\\ \mathrm{vec}\boldsymbol{S}-\mathrm{vec}\boldsymbol{\Sigma}\end{bmatrix} \tag{13}$$

$$\xrightarrow{d} N\left[\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{0}\\ \boldsymbol{0} & (\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\kappa\mathrm{vec}\boldsymbol{\Sigma}(\mathrm{vec}\boldsymbol{\Sigma})'\end{pmatrix}\right].$$

这个定理可由独立同分布随机向量的中心极限定理推出 (四阶矩有限). 这个定理形成了大样本推断的基础.

3.6.2 峰度参数的估计

为了将正态分布的大样本分布理论应用到椭球等高分布的推断问题上, 我们需要知道或估计峰度参数 κ. 注意

$$\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})]^2=\left(\frac{p}{\mathrm{E}(R^2)}\right)^2\mathrm{E}(\boldsymbol{Y}'\boldsymbol{Y})^2 \tag{14}$$
$$=\frac{p^2\mathrm{E}(R^4)}{(\mathrm{E}(R^2))^2}=p(p+2)(1+\kappa).$$

因为 $\bar{\boldsymbol{x}}\xrightarrow{p}\boldsymbol{\mu}$ 和 $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Sigma}$, 则

$$\frac{1}{N}\sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})]^2\xrightarrow{p}p(p+2)(1+\kappa). \tag{15}$$

κ 的一个一致估计是

$$\hat{\kappa}=\frac{1}{p(p+2)}\frac{1}{N}\sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})]^2-1. \tag{16}$$

Mardia(1970) 提出用 M 形成一个 κ 的一致估计.

3.6.3 极大似然估计

我们考虑用 $\boldsymbol{S}$ 作为 $\boldsymbol{\Sigma}=(\mathrm{E}(R^2/p))\boldsymbol{\Lambda}$ 的一个估计. 当母分布是正态的时候, $\boldsymbol{S}$ 是充分统计量且对变换不变, 因而是有效无偏估计. 现在我们来研究其他估计.

我们首先考虑当密度 $g(\cdot)$ 的形式已知时, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Lambda}$ 的极大似然估计. 其似然函数的对数是

$$\ln L=-\frac{N}{2}\ln|\boldsymbol{\Lambda}|+\sum_{\alpha=1}^{N}\ln g[(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})]. \tag{17}$$

$\ln L$ 关于 $\boldsymbol{\mu}$ 的分量的导数是

$$\frac{\partial\ln L}{\partial\boldsymbol{\mu}}=-2\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})]}{g[(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})]}\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}). \tag{18}$$

令导数向量为 $\mathbf{0}$, 则有方程

$$\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}\boldsymbol{x}_\alpha=\hat{\boldsymbol{\mu}}\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}. \tag{19}$$

令 $\ln L$ 关于 $\boldsymbol{\Lambda}^{-1}$ 的元素的导数等于 0, 得到

$$\hat{\boldsymbol{\Lambda}}=-\frac{2}{N}\sum_{\alpha=1}^{N}\frac{g'[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}{g[(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Lambda}}^{-1}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})]}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'. \tag{20}$$

估计 $\hat{\boldsymbol{\Lambda}}$ 是秩为 1 的矩阵 $(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\mu}})'$ 的一种加权平均. 在正态情形下, 权重为 $1/N$. 在大多数情形下, (19) 和 (20) 不能显示求解, 但是其解可以用迭代方法逼近.

$\sqrt{N}(\mathrm{vec}\hat{\boldsymbol{\Lambda}}-\mathrm{vec}\boldsymbol{\Lambda})$ 的极限正态分布的协方差阵是

$$\mathrm{C}(\mathrm{vec}\hat{\boldsymbol{\Lambda}})=\sigma_{1g}(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Lambda}\otimes\boldsymbol{\Lambda})+\sigma_{2g}\mathrm{vec}\boldsymbol{\Lambda}(\mathrm{vec}\boldsymbol{\Lambda})', \tag{21}$$

其中

$$\sigma_{1g}=\frac{p(p+2)}{4\mathrm{E}\left[\frac{g'(R^2)}{g(R^2)}R^2\right]^2}, \tag{22}$$

$$\sigma_{2g} = \frac{2\sigma_{1g}(1-\sigma_{1g})}{2+p(1-\sigma_{1g})}. \tag{23}$$

见 Tyler(1982).

3.6.4 椭球等高矩阵分布

设

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1' \\ \vdots \\ \boldsymbol{y}_N' \end{bmatrix} \tag{24}$$

是一个 $N \times p$ 随机矩阵, 其密度是 $g(\boldsymbol{Y}'\boldsymbol{Y}) = g(\sum_{\alpha=1}^N \boldsymbol{y}_\alpha \boldsymbol{y}_\alpha')$. 注意, 密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$ 关于正交变换 $\boldsymbol{Y}^* = \boldsymbol{O}_N\boldsymbol{Y}$ 是不变的. 这样的密度称为*左球面矩阵*密度. 一个例子是来自 $N(\boldsymbol{0}, \boldsymbol{I}_p)$ 的容量为 N 的样本的密度

$$g(\boldsymbol{Y}'\boldsymbol{Y}) = \frac{1}{(2\pi)^{\frac{1}{2}pN}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{Y}'\boldsymbol{Y}}. \tag{25}$$

在这个例子中, $\boldsymbol{Y}$ 也是*右球面的*: $\boldsymbol{Y}\boldsymbol{O}_p \stackrel{d}{=} \boldsymbol{Y}$. 当 $\boldsymbol{Y}$ 既是左球面的又是右球面的时, 我们称它为*球面的*. 进而, 如果 $\boldsymbol{Y}$ 有密度 (25), 则 $\mathrm{vec}\boldsymbol{Y}$ 是球面的. 一般地, 如果 $\boldsymbol{Y}$ 的密度形式是

$$\begin{aligned} g(\mathrm{tr}\boldsymbol{Y}'\boldsymbol{Y}) &= g\left(\sum_{\alpha=1}^N \sum_{i=1}^p y_{i\alpha}^2\right) = g(\mathrm{tr}\boldsymbol{Y}\boldsymbol{Y}') \\ &= g[(\mathrm{vec}\boldsymbol{Y})'\mathrm{vec}\boldsymbol{Y}] = g[(\mathrm{vec}\boldsymbol{Y}')'\mathrm{vec}\boldsymbol{Y}'], \end{aligned} \tag{26}$$

我们称这个模型为*向量–球面的*. 定义

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{C}' + \boldsymbol{\varepsilon}_N\boldsymbol{\mu}', \tag{27}$$

其中 $\boldsymbol{C}'\boldsymbol{\Lambda}^{-1}\boldsymbol{C} = \boldsymbol{I}_p$ 和 $\boldsymbol{\varepsilon}_N' = (1, \cdots, 1)$. 因为 (27) 等价于 $\boldsymbol{Y} = (\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\mu}')(\boldsymbol{C}')^{-1}$ 且 $(\boldsymbol{C}')^{-1}\boldsymbol{C}^{-1} = \boldsymbol{\Lambda}^{-1}$, 则矩阵 $\boldsymbol{X}$ 有密度

$$\begin{aligned} &|\boldsymbol{\Lambda}|^{-N/2} g[\mathrm{tr}(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\mu}')\boldsymbol{\Lambda}^{-1}(\boldsymbol{X} - \boldsymbol{\varepsilon}_N\boldsymbol{\mu}')'] \\ &= |\boldsymbol{\Lambda}|^{-N/2} g\left[\sum_{\alpha=1}^N (\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right]. \end{aligned} \tag{28}$$

从 (26), 我们推出 $\mathrm{vec}\boldsymbol{Y}$ 有表示

$$\mathrm{vec}\boldsymbol{Y} \stackrel{d}{=} R\,\mathrm{vec}\boldsymbol{U} \tag{29}$$

其中 $w = R^2$ 有密度

$$\frac{\pi^{\frac{1}{2}Np}}{\Gamma(Np/2)} w^{\frac{1}{2}Np-1} g(w), \tag{30}$$

$\mathrm{vec}\boldsymbol{U}$ 在 $\sum_{\alpha=1}^N \sum_{i=1}^p u_{i\alpha}^2 = 1$ 上服从均匀分布, 且 R 和 $\mathrm{vec}\boldsymbol{U}$ 独立. $\mathrm{vec}\boldsymbol{Y}$ 的协方差阵是

$$\mathrm{E}[\mathrm{vec}\boldsymbol{Y}(\mathrm{vec}\boldsymbol{Y})'] = \frac{\mathrm{E}(R^2)}{Np}\boldsymbol{I}_{Np} = \frac{\mathrm{E}(R^2)}{Np}(\boldsymbol{I}_p \otimes \boldsymbol{I}_N). \tag{31}$$

因为对任意合适的矩阵 $\boldsymbol{F},\boldsymbol{G}$ 和 $\boldsymbol{H}$, 有 $\boldsymbol{FGH} = (\boldsymbol{H}'\otimes\boldsymbol{F})\mathrm{vec}\boldsymbol{G}$, 我们可以将 (27) 写为

$$\mathrm{vec}\boldsymbol{X} = (\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{Y} + \boldsymbol{\mu}\otimes\boldsymbol{\varepsilon}_N. \tag{32}$$

因此

$$\mathrm{Eov}(\mathrm{vec}\boldsymbol{X}) = \boldsymbol{\mu}\otimes\boldsymbol{\varepsilon}_N, \tag{33}$$

$$\mathrm{Cov}(\mathrm{vec}\boldsymbol{X}) = (\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{C}(\mathrm{vec}\boldsymbol{Y})(\boldsymbol{C}'\otimes\boldsymbol{I}_N) = \frac{\mathrm{E}(R^2)}{Np}\boldsymbol{\Lambda}\otimes\boldsymbol{I}_N, \tag{34}$$

$$\mathrm{E}(\boldsymbol{X}\text{ 的行}) = \boldsymbol{\mu}', \tag{35}$$

$$\mathrm{Cov}(\boldsymbol{X}'\text{的行}) = \frac{\mathrm{E}(R^2)}{Np}\boldsymbol{\Lambda}. \tag{36}$$

$\boldsymbol{X}$ 的行是不相关的 (虽然没必要独立). 从 (32) 我们得到

$$\mathrm{vec}\boldsymbol{X} \overset{d}{=} \boldsymbol{R}(\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{U} + \boldsymbol{\mu}\otimes\boldsymbol{\varepsilon}_N, \tag{37}$$

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{RUC}' + \boldsymbol{\varepsilon}_N\boldsymbol{\mu}'. \tag{38}$$

因为 $\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\mu}' = (\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}') + \boldsymbol{\varepsilon}_N(\bar{\boldsymbol{X}}-\boldsymbol{\mu})'$ 且 $\boldsymbol{\varepsilon}_N'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}') = \boldsymbol{0}$, 我们可将 $\boldsymbol{X}$ 的密度写为

$$|\boldsymbol{\Lambda}|^{-N/2}g[\mathrm{tr}\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}') + N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})], \tag{39}$$

其中 $\bar{\boldsymbol{x}} = (1/N)\boldsymbol{X}'\boldsymbol{\varepsilon}_N$. 这说明和正态分布一样, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Lambda}$ 的一个充分统计量集合是 $\bar{\boldsymbol{x}}$ 和 $n\boldsymbol{S} = (\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')$, 极大似然估计可以由下面的定理得到, 它也将被用在后面其他模型中.

定理 3.6.3 假设 m 维向量 $\boldsymbol{Z}$ 有密度 $|\boldsymbol{\Phi}|^{-\frac{1}{2}}h[(\boldsymbol{z}-\boldsymbol{\nu})'\boldsymbol{\Phi}^{-1}(\boldsymbol{z}-\boldsymbol{\nu})]$, 其中 $w^{\frac{1}{2}m}h(w)$ 在 w_h 处有一个有限正极大值, $\boldsymbol{\Phi}$ 是正定矩阵. 设 Ω 是 $(\boldsymbol{\nu},\boldsymbol{\Phi})$ 空间里的一个集合, 使得对所有 $c>0$, 如果 $(\boldsymbol{\nu},\boldsymbol{\Phi})\in\Omega$ 则 $(\boldsymbol{\nu},c\boldsymbol{\Phi})\in\Omega$. 假设基于一个观测 $\boldsymbol{z}$, 当 $h(w) = \mathrm{const}\mathrm{e}^{-\frac{1}{2}w}$(即 $\boldsymbol{Z}$ 服从正态分布), 极大似然估计 $(\bar{\boldsymbol{\nu}},\bar{\boldsymbol{\Phi}})\in\Omega$ 存在且唯一, 其中 $\bar{\boldsymbol{\Phi}}$ 以概率 1 是正定的. 则对任意函数 $h(\cdot)$, $(\boldsymbol{\nu},\boldsymbol{\Phi})$ 的极大似然估计是

$$\hat{\boldsymbol{\nu}} = \bar{\boldsymbol{\nu}}, \qquad \hat{\boldsymbol{\Phi}} = \frac{m}{w_h}\bar{\boldsymbol{\Phi}}, \tag{40}$$

并且极大似然是 $|\hat{\boldsymbol{\Phi}}|^{-\frac{1}{2}}h(w_h)$ [Anderson, Fang, and Hsu (1986)].

证明 设 $\boldsymbol{\Psi} = |\boldsymbol{\Phi}|^{-1/m}\boldsymbol{\Phi}$ 且

$$d = (\boldsymbol{z}-\boldsymbol{\nu})'\boldsymbol{\Phi}^{-1}(\boldsymbol{z}-\boldsymbol{\nu}) = \frac{(\boldsymbol{z}-\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{z}-\boldsymbol{\nu})}{|\boldsymbol{\Phi}|^{1/m}}. \tag{41}$$

则 $(\boldsymbol{\nu},\boldsymbol{\Phi})\in\Omega$ 且 $|\boldsymbol{\Psi}|=1$. 似然是

$$[(\boldsymbol{z}-\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{z}-\boldsymbol{\nu})]^{-\frac{1}{2}m}d^{\frac{1}{2}m}h(d). \tag{42}$$

在正态条件下 $h(d) = (2\pi)^{-\frac{1}{2}m}\mathrm{e}^{-\frac{1}{2}d}$, 且 (42) 的极大值点在 $\boldsymbol{\nu}=\bar{\boldsymbol{\nu}}, \boldsymbol{\Psi}=\bar{\boldsymbol{\Psi}}=|\bar{\boldsymbol{\Phi}}|^{-1/m}\bar{\boldsymbol{\Phi}}$, $d=m$. 对任意 $h(\cdot)$, (42) 的极大值点在 $\hat{\boldsymbol{\nu}}=\bar{\boldsymbol{\nu}}, \hat{\boldsymbol{B}}=\bar{\boldsymbol{B}}$, $\hat{d}=w_h$. 则 $\boldsymbol{\Phi}$ 的极大似然估计是

$$\hat{\boldsymbol{\Phi}} = |\hat{\boldsymbol{\Phi}}|^{1/m}\hat{\boldsymbol{\Psi}} = \frac{|\hat{\boldsymbol{\Phi}}|^{1/m}}{|\bar{\boldsymbol{\Phi}}|^{1/m}}\bar{\boldsymbol{\Phi}}. \tag{43}$$

则利用 (41), 由 (43) 可得 (40). ■

定理 3.6.4 设 $\boldsymbol{X}(N\times p)$ 有密度 (28), 其中 $w^{\frac{1}{2}Np}g(w)$ 在 w_g 处取有限正极大值. 则 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Lambda}$ 的极大似然估计是

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}, \qquad \hat{\boldsymbol{\Lambda}} = \frac{Np}{w_g}\boldsymbol{A}, \tag{44}$$

其中 $\boldsymbol{A} = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$.

推论 3.6.1 设 $\boldsymbol{X}(N\times p)$ 有密度 (28). 则 $\boldsymbol{\nu}$, $(\lambda_{11},\cdots,\lambda_{pp})$ 和 $\rho_{ij}, i,j=1,\cdots,p$ 的极大似然估计是 $\bar{x}$, $(p/w_g)(a_{11},\cdots,a_{pp})$ 和 $a_{ij}/\sqrt{a_{ii}a_{jj}}, i,j=1,\cdots,p$.

证明 推论 3.6.1 可由定理 3.6.3 和推论 3.2.1 得出. ■

定理 3.6.5 设 $\boldsymbol{f}(\boldsymbol{X})$ 是 $\boldsymbol{X}(N\times p)$ 的向量值函数, 满足: 对所有的 $\boldsymbol{\nu}$, 有

$$\boldsymbol{f}(\boldsymbol{X} + \boldsymbol{\varepsilon}_N\boldsymbol{\nu}') = \boldsymbol{f}(\boldsymbol{X}); \tag{45}$$

对所有的 c, 有

$$\boldsymbol{f}(c\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X}). \tag{46}$$

则当 $\boldsymbol{X}$ 有任意密度 (28) 时和当 $\boldsymbol{X}$ 有正态密度 (28) 时, $\boldsymbol{f}(\boldsymbol{X})$ 的分布是一样的.

证明 把表示 (27) 代入 $\boldsymbol{f}(\boldsymbol{X})$, 并由 (45) 得到

$$\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{Y}\boldsymbol{C}' + \boldsymbol{\varepsilon}_N\boldsymbol{\mu}') = \boldsymbol{f}(\boldsymbol{Y}\boldsymbol{C}'). \tag{47}$$

设 $\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{h}(\mathrm{vec}\boldsymbol{X})$. 则由 (46), $\boldsymbol{h}(c\boldsymbol{X}) = \boldsymbol{h}(\boldsymbol{X})$ 且

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{Y}\boldsymbol{C}') &= \boldsymbol{h}[(\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{Y}] = \boldsymbol{h}[R(\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{U}] \\ &= \boldsymbol{h}[(\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{U}].\end{aligned} \tag{48}$$

■

对所有的 $g(\cdot)$, 任何满足 (45) 和 (46) 的统计量都有相同的分布. 因此, 如果在正态情形下这个分布已知, 那么这个分布对所有椭球等高分布都是有效的.

任何充分统计量集合的函数, 如果它是变换–不变的, 即满足 (45), 则它是 $\boldsymbol{S}$ 的一个函数. 因此关于 $\boldsymbol{\Sigma}$ 的推断可以建立在 $\boldsymbol{S}$ 的基础上.

推论 3.6.2 设 $\boldsymbol{f}(\boldsymbol{X})$ 是 $\boldsymbol{X}$ $(N\times p)$ 的向量值函数, 对所有 $c>0$ 满足 (46). 则当 $\boldsymbol{X}$ 有 $\boldsymbol{\mu}=\boldsymbol{0}$ 的任意密度 (28) 时和 $\boldsymbol{X}$ 有 $\boldsymbol{\mu}=\boldsymbol{0}$ 的正态密度 (28) 时, $\boldsymbol{f}(\boldsymbol{X})$ 的分布是一样的.

Fang and Zhang (1990) 以定理 2.5.8 给出了这个推论.

习 题

3.1 (3.2 节) 由摘自 Frets (1921) 的表 3.3 给出的数据求 $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$ 和 $(\hat{\rho}_{ij})$.

表 3.3[①]　兄弟的头长和头宽

头长 长子 x_1	头宽 长子 x_2	头长 次子 x_3	头宽 次子 x_4
191	155	179	145
195	149	201	152
181	148	185	149
183	153	188	149
176	144	171	142
208	157	192	152
189	150	190	149
197	159	189	152
188	152	197	159
192	150	187	151
179	158	186	148
183	147	174	147
174	150	185	152
190	159	195	157
188	151	187	158
163	137	161	130
195	155	183	158
186	153	173	148
181	145	182	146
175	140	165	137
192	154	185	152
174	143	178	147
176	139	176	143
197	167	200	158
190	163	187	150

3.2　(3.2 节)　证明 (21) 的数值结果.

3.3　(3.2 节)　由下面几对观测, 计算 $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, $\boldsymbol{S}$ 和 $\hat{\rho}$: (34, 55), (12, 29), (33, 75), (44, 89), (89, 62), (59, 69), (50, 41), (88, 67). 描出这些观测值.

3.4　(3.2 节)　利用事实 $|\boldsymbol{C}^*| = \prod \lambda_i$, $\mathrm{tr}\boldsymbol{C}^* = \sum \lambda_i$, 和 $\boldsymbol{C}^* = \boldsymbol{I}$ 如果 $\lambda_1 = \cdots = \lambda_p = 1$, 其中 $\lambda_1, \cdots, \lambda_p$ 是 $\boldsymbol{C}^*$ 的特征根, 证明引理 3.2.2. [提示: 利用 (12) 里给出的 f.]

3.5　(3.2 节)　设 x_1 是一只猫的体重 (单位千克), x_2 是其心脏的重量 (单位克). [数据来自

① 这些数据曾被本书第 1 版用作例子, 它们来自 Rao (1952), 第 245 页. Izenman (1980) 已经指出有些数明显错误地从 Frets (1921) 复制出并更正了它们 (第 579 页).

Fisher (1947b).]

(a) 一个有 47 只母猫的样本的相关数据是

$$\sum \boldsymbol{x}_\alpha = \begin{pmatrix} 110.9 \\ 432.5 \end{pmatrix}, \qquad \sum \boldsymbol{x}_\alpha \boldsymbol{x}'_\alpha = \begin{pmatrix} 265.13 & 1029.62 \\ 1029.62 & 4064.71 \end{pmatrix}.$$

求 $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, $\boldsymbol{S}$ 和 $\hat{\rho}$.

(b) 一个有 97 只公猫的样本的相关数据是

$$\sum \boldsymbol{x}_\alpha = \begin{pmatrix} 281.3 \\ 1098.3 \end{pmatrix}, \qquad \sum \boldsymbol{x}_\alpha \boldsymbol{x}'_\alpha = \begin{pmatrix} 836.75 & 3275.55 \\ 3275.55 & 13056.17 \end{pmatrix}.$$

求 $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$, $\boldsymbol{S}$ 和 $\hat{\rho}$.

3.6 从表 3.4 求山鸢尾的 $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$ 和 $(\hat{\rho}_{ij})$, 表 3.4 选自 Edgar Anderson 著名的鸢尾属植物数据 [Fisher (1936)].

表 3.4 三种鸢尾属植物的四个测量 (单位厘米)

山鸢尾				变色鸢尾				维吉尼亚鸢尾			
花萼长度	花萼宽度	花瓣长度	花瓣宽度	花萼长度	花萼宽度	花瓣长度	花瓣宽度	花萼长度	花萼宽度	花瓣长度	花瓣宽度
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0

(续)

山鸢尾				变色鸢尾				维吉尼亚鸢尾			
花萼长度	花萼宽度	花瓣长度	花瓣宽度	花萼长度	花萼宽度	花瓣长度	花瓣宽度	花萼长度	花萼宽度	花瓣长度	花瓣宽度
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
4.6	3.2	7.4	0.2	6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
530	3.3	1.4	0.2	5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8

3.7　(3.2 节)　**样本相关系数的不变性.** 证明一个二元样本的 r_{12} 在平移和比例变换 ($x_{i\alpha}^* = b_i x_{i\alpha} + c_i, b_i > 0, i = 1, 2, \alpha = 1, \cdots, N$) 下, 是充分统计量 $\bar{x}$ 和 S 的不变特征, 并且每一个 $\bar{x}$ 和 S 的不变函数都是 r_{12} 的函数. [提示: 见定理 2.3.2.]

3.8　(3.2 节)　用归纳法证明引理 3.2.2. [提示: 设 $\boldsymbol{H}_1 = h_{11}$,

$$\boldsymbol{H}_i = \begin{pmatrix} \boldsymbol{H}_{i-1} & \boldsymbol{h}_{(i)} \\ \boldsymbol{h}'_{(1)} & h_{ii} \end{pmatrix}, \quad i = 2, \cdots, p,$$

并利用习题 2.36.]

3.9 (3.2 节) 证明

$$\frac{1}{N(N-1)}\sum_{\alpha<\beta}(\boldsymbol{x}_\alpha-\boldsymbol{x}_\beta)(\boldsymbol{x}_\alpha-\boldsymbol{x}_\beta)'=\frac{1}{N-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'.$$

(注意: 当 $p=1$, 左端是观测的平均平方差.)

3.10 (3.2 节) **当 $\boldsymbol{\mu}$ 已知时 $\boldsymbol{\Sigma}$ 的估计.** 证明: 如果 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本且 $\boldsymbol{\mu}$ 已知, 则 $(1/N)\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'$ 是 $\boldsymbol{\Sigma}$ 的极大似然估计.

3.11 (3.2 节) **复正态分布的参数估计.** 设 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N$ 是来自均值为 $\boldsymbol{\theta}$, 协方差阵为 $\boldsymbol{P}$ 的复正态分布的 N 个观测. (见习题 2.64.)

(a) 证明 $\boldsymbol{\theta}$ 和 $\boldsymbol{P}$ 的极大似然估计是

$$\hat{\boldsymbol{\theta}}=\bar{\boldsymbol{z}}=\frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha,\quad \hat{\boldsymbol{P}}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{z}_\alpha-\bar{\boldsymbol{z}})(\boldsymbol{z}_\alpha-\bar{\boldsymbol{z}})^*.$$

(b) 证明 $\bar{\boldsymbol{z}}$ 服从均值为 $\boldsymbol{\theta}$, 协方差阵为 $(1/N)\boldsymbol{P}$ 的复正态分布.

(c) 证明 $\bar{\boldsymbol{z}}$ 和 $\hat{\boldsymbol{P}}$ 独立, 且 $N\hat{\boldsymbol{P}}$ 与 $\sum_{\alpha=1}^{n}\boldsymbol{W}_\alpha\boldsymbol{W}_\alpha^*$ 同分布, 其中 $\boldsymbol{W}_1,\cdots,\boldsymbol{W}_n$ 独立且都服从均值为 $\boldsymbol{\theta}$, 协方差阵为 $\boldsymbol{P}$ 的复正态分布, 这里 $n=N-1$.

3.12 (3.2 节) 用引理 3.2.3 证明引理 3.2.2, 并且通过令 $N\ln|\boldsymbol{C}|-\mathrm{tr}\boldsymbol{CD}$ 关于 $\boldsymbol{C}=\boldsymbol{\Sigma}^{-1}$ 的元素的导数为 0, 证明其在 $\boldsymbol{C}=N\boldsymbol{D}^{-1}$ 处有极大值. 证明当 $\boldsymbol{C}$ 趋于一个奇异矩阵时或者 $\boldsymbol{C}$ 的一个或多个元素趋于 ∞ 和 (或)$-\infty$ (非对角元) 时, 这个 $\boldsymbol{C}$ 的函数趋于 $-\infty$. 为了证明后者, 可利用 (13) 的等价性.

3.13 (3.3 节) 设 $\boldsymbol{X}_\alpha$ 服从分布 $N(\boldsymbol{\gamma}c_\alpha,\boldsymbol{\Sigma}),\alpha=1,\cdots,N$, 其中 $\sum c_\alpha^2>0$. 证明 $\boldsymbol{g}=(1/\sum c_\alpha^2)\sum c_\alpha\boldsymbol{X}_\alpha$ 服从分布 $N[\boldsymbol{\gamma},(1/\sum c_\alpha^2)\boldsymbol{\Sigma}]$. 证明 $\boldsymbol{E}=\sum_\alpha(\boldsymbol{X}_\alpha-\boldsymbol{g}c_\alpha)(\boldsymbol{X}_\alpha-\boldsymbol{g}c_\alpha)'$ 独立同分布于 $\sum_{\alpha=1}^{N-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_N$ 独立且都服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. [**提示:** 设 $\boldsymbol{Z}_\alpha=\sum b_{\alpha\beta}\boldsymbol{X}_\beta$, 其中 $b_{N\beta}=c_\beta/\sqrt{\sum c_\alpha^2}$ 且 $\boldsymbol{B}$ 是正交的.]

3.14 (3.3 节) 给定 α, 证明 (19) 中检验的功效是一个只与 p 和 $[N_1N_2/(N_1+N_2)](\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$ 有关的函数.

3.15 (3.3 节) **均值的效率.** 证明 $\bar{\boldsymbol{x}}$ 对估计 $\boldsymbol{\mu}$ 是有效的.

3.16 (3.3 节) 证明对于估计$\boldsymbol{\mu}$和 $\boldsymbol{\Sigma}$, $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 有效率 $[(N-1)/N]^{p(p+1)/2}$.

3.17 (3.2 节) 证明对于由 (4) 定义的 $\boldsymbol{A}$, 当 $N>p$ 时, $\Pr\{|\boldsymbol{A}|=0\}=0$. [**提示:** 讨论如果 $\boldsymbol{Z}_p^*=(\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_p)$, 则 $|\boldsymbol{Z}_p^*|\neq 0$ 意味着 $\boldsymbol{A}=\boldsymbol{Z}_p^*\boldsymbol{Z}_p^{*\prime}+\sum_{\alpha=p+1}^{N-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 是正定的. 用归纳法证明 $\Pr\{|\boldsymbol{Z}_j^*|=Z_{jj}|\boldsymbol{Z}_{j-1}^*|+\sum_{i=1}^{j-1}Z_{ij}\mathrm{cof}(Z_{ij})=0\}=0$, $j=2,\cdots,p$.]

3.18 (3.4 节) 证明

$$\boldsymbol{I}-\boldsymbol{\Phi}(\boldsymbol{\Phi}+\boldsymbol{\Sigma})^{-1}=\boldsymbol{\Sigma}(\boldsymbol{\Phi}+\boldsymbol{\Sigma})^{-1},$$

$$\boldsymbol{\Phi}-\boldsymbol{\Phi}(\boldsymbol{\Phi}+\boldsymbol{\Sigma})^{-1}\boldsymbol{\Phi}=(\boldsymbol{\Phi}^{-1}+\boldsymbol{\Sigma}^{-1})^{-1}.$$

3.19 (3.4 节) 证明当 $\boldsymbol{\mu}$ 已知时, $(1/N)\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'$ 是 $\boldsymbol{\Sigma}$ 的一个无偏估计.

3.20 (3.4 节) 证明

$$\boldsymbol{\Phi}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{x}+\frac{1}{N}\boldsymbol{\Sigma}\left(\boldsymbol{\Phi}+\frac{1}{N}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\nu}=(\boldsymbol{\Phi}^{-1}+N\boldsymbol{\Sigma}^{-1})^{-1}(N\boldsymbol{\Sigma}^{-1}\boldsymbol{x}+\boldsymbol{\Phi}^{-1}\boldsymbol{\nu}).$$

3.21 (3.5 节) 用分部积分的方法论证引理 3.5.1.

3.22 (3.5 节) 证明

$$\int_\theta^\infty \int_y^\infty \left| f'(y)(x-\theta)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2} \right| \mathrm{d}x\mathrm{d}y = \int_\theta^\infty |f'(y)| \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(y-\theta)^2}\mathrm{d}y,$$

$$\int_{-\infty}^\theta \int_{-\infty}^y \left| f'(y)(\theta-x)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\theta)^2} \right| \mathrm{d}x\mathrm{d}y = \int_{-\infty}^\theta |f'(y)| \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(y-\theta)^2}\mathrm{d}y.$$

3.23 设 $\boldsymbol{Z}(k) = (Z_{ij}(k))$ 是一个随机矩阵序列, 其中 $i = 1,\cdots,p$, $j = 1,\cdots,q$ 和 $k = 1,2,\cdots$. 设矩阵 $\boldsymbol{A}$ 的一个范数是 $N_1(\boldsymbol{A}) = \max_{i,j}\mathrm{mod}(a_{ij})$, 另一个范数是 $N_2(\boldsymbol{A}) = \sum_{i,j} a_{ij}^2 = \mathrm{tr}\boldsymbol{A}\boldsymbol{A}'$. 一些定义 $\boldsymbol{Z}(k)$ 随机收敛到 $\boldsymbol{B}$ $(p\times q)$ 的方法如下.

(a) $N_1(\boldsymbol{Z}(k)-\boldsymbol{B})$ 随机收敛到 0.

(b) $N_2(\boldsymbol{Z}(k)-\boldsymbol{B})$ 随机收敛到 0.

(c) $Z_{ij}(k)-b_{ij}$ 随机收敛到 0, $i=1,\cdots,p$, $j=1,\cdots,q$.

证明这三个定义是等价的. 注意 $X(k)$ 随机收敛到 a 的定义是, 对任意正数 δ 和 ε, 我们能找到足够大的 K, 使得对 $k>K$, 有

$$\Pr\{|X(k)-a|<\delta\} > 1-\varepsilon.$$

3.24 (3.2 节) **线性结构的协方差阵** [Anderson (1969)]. 设

$$\boldsymbol{\Sigma} = \sum_{g=0}^{q} \sigma_g \boldsymbol{G}_g, \tag{i}$$

其中 $\boldsymbol{G}_0,\cdots,\boldsymbol{G}_q$ 是给定的对称矩阵, 满足至少存在一个 $(q+1)$ 多元组, $\sigma_0,\sigma_1,\cdots,\sigma_q$ 使得 (i) 是正定的. 证明基于 N 个观测的似然方程是

$$-\frac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}_g + \frac{1}{2}\mathrm{tr}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}\boldsymbol{G}_g\boldsymbol{\Sigma}^{-1} = 0, \quad g=0,1,\cdots,q. \tag{ii}$$

证明一个迭代 (得分) 方法可以基于

$$\sum_{h=0}^{q} \mathrm{tr}\hat{\boldsymbol{\Sigma}}_{i-1}^{-1}\boldsymbol{G}_g\hat{\boldsymbol{\Sigma}}_{i-1}^{-1}\boldsymbol{G}_h\hat{\sigma}_h^{(i)} = \frac{1}{N}\mathrm{tr}\hat{\boldsymbol{\Sigma}}_{i-1}^{-1}\boldsymbol{G}_g\hat{\boldsymbol{\Sigma}}_{i-1}^{-1}\boldsymbol{A}, \quad g=0,1,\cdots,q, \tag{iii}$$

其中 $\hat{\boldsymbol{\Sigma}}_{i-1}^{-1}\sum_{g=0}^{q}\hat{\sigma}_g^{(i-1)}\boldsymbol{G}_g$.

第 4 章　样本相关系数的分布和利用

4.1 引　言

第 2 章介绍了多元正态分布, 提出了一个关于两正态变量相依性的测度, 即相关系数 $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. 在给定 $X_{q+1} = x_{q+1}, \cdots, X_p = x_p$ 时 $X_1, \cdots, X_q$ 的条件分布里, 偏相关系数 $\rho_{ij \cdot q+1,\cdots,p}$ 度量了 X_i 和 X_j 的相依性. 讨论的第三种相关性是多重相关, 它度量了一个变量和一组其他变量之间的关系. 本章我们处理这些量的样本等价量, 它们是那些总体量的点估计. 求出样本相关系数的分布. 讨论假设检验和置信区间.

在联合正态分布情形下, 这些相关系数是相依性的自然测度. 它们是总体除了位置 (均值) 和散度 (标准差) 参数外仅有的参数. 在样本中, 相关系数是总体相关系数的合理估计. 因为样本均值和标准差是位置和散度的估计, 所以样本相关系数 (即标准化样本二阶矩) 给出了关于总体相关系数的所有可能信息. 样本相关系数是关于位置和散度变换不变的充分统计量的函数; 总体相关系数是关于这些变换不变的参数的函数.

在*回归理论*或最小二乘中, 有一个变量视为随机或*相依*的, 其他变量视为固定或*自变*的. 在相关性理论中, 我们视一些变量为随机的并对称地处理它们. 若从一个联合正态分布开始, 固定除一个变量外的所有变量, 则我们可以得到最小二乘模型, 因为随机变量在其条件分布下的期望值是那些固定变量的线性函数. 最小二乘中得到的样本回归系数是样本方差和相关系数.

在独立性检验中, 我们将看到, 不论哪种情况 (即在联合正态分布或最小二乘的条件分布中), 都能达到相同的检验. 在原假设下的概率理论是一样的. 当原假设不真时, 这两种情况检验判据的分布是不同的. 如果所有变量都是随机的, 那么可利用这里介绍的相关理论; 如果只有一个变量是随机的, 那么可以利用最小二乘理论 (第 8 章会在一些普遍情况下考虑它).

4.2 节将推导样本相关系数的分布, 首先是当相应总体相关系数为 0 (两个正态变量独立) 时的情况, 其次是当总体相关系数为任意值时的情况. Fisher z 变换可以产生一个有用的近似正态分布. 我们在 4.2 节将会讨论精确或近似置信区间. 4.3 节我们讨论关于偏相关系数的一些方法, 即条件正态分布中相关性的相同问题. 4.4 节研究样本多重相关系数的分布和其他性质. 4.5 节得到了椭球等高分布的渐近分布, 并且给出了这类分布的一个随机表示.

4.2 二元变量样本的相关系数

4.2.1 当总体相关系数为零时的分布, 相关性不足的假设检验

在 3.2 节中证明了, 如果样本 (p 维向量) $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是从正态分布得到的, 则 X_i 和 X_j (随机向量 $\boldsymbol{X}$ 的两个分量) 相关系数的极大似然估计是

$$r_{ij} = \frac{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)}{\sqrt{\sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)^2}\sqrt{\sum_{\alpha=1}^{N}(x_{j\alpha} - \bar{x}_j)^2}}, \tag{1}$$

其中 $x_{i\alpha}$ 是 $\boldsymbol{x}_\alpha$ 的第 i 个分量且

$$\bar{x}_i = \frac{1}{N}\sum_{\alpha=1}^{N} x_{i\alpha}. \tag{2}$$

本节我们将求当 X_i 和 X_j 的总体相关系数为 0 时 r_{ij} 的分布, 并将看到怎样利用样本相关系数来检验总体相关系数为 0 的假设.

为了方便, 我们考虑 r_{12}, 每个 r_{ij} 都有相同的理论. 因为 r_{12} 只依赖每个 $\boldsymbol{x}_\alpha$ 的前两个分量, 因此为了求 r_{12} 的分布, 我们只需要考虑 $(x_{11}, x_{21}), (x_{12}, x_{22}), \cdots, (x_{1N}, x_{2N})$. 我们可以把这里考虑的问题从二元正态分布的角度重新表示一下. 设 $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_N^*$ 是来自

$$N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{pmatrix}\right] \tag{3}$$

的观测向量. 我们将考虑

$$r = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}, \tag{4}$$

其中

$$a_{ij} = \sum_{\alpha=1}^{N}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \quad i, j = 1, 2, \tag{5}$$

且 $\bar{x}_i$ 由 (2) 定义, $x_{i\alpha}$ 是 $\boldsymbol{x}_\alpha^*$ 的第 i 个分量.

从 3.3 节中我们可以看到 a_{11}, a_{12} 和 a_{22} 的分布如同

$$a_{ij} = \sum_{\alpha=1}^{n} z_{i\alpha} z_{j\alpha}, \quad i, j = 1, 2, \tag{6}$$

其中 $n = N - 1$, $(z_{1\alpha}, z_{2\alpha})$ 服从分布

$$N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{pmatrix}\right], \tag{7}$$

且 $(z_{11}, z_{21}), \cdots, (z_{1N}, z_{2N})$ 相互独立分布.

定义 n 维向量 $\boldsymbol{v}_i = (z_{i1}, \cdots, z_{in})'$, $i = 1, 2$. 这两个向量可以表示在一个 n 维空间里, 见图 4.1. 它们的相关系数是 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 的夹角 θ 的余弦. (见 3.2 节.)

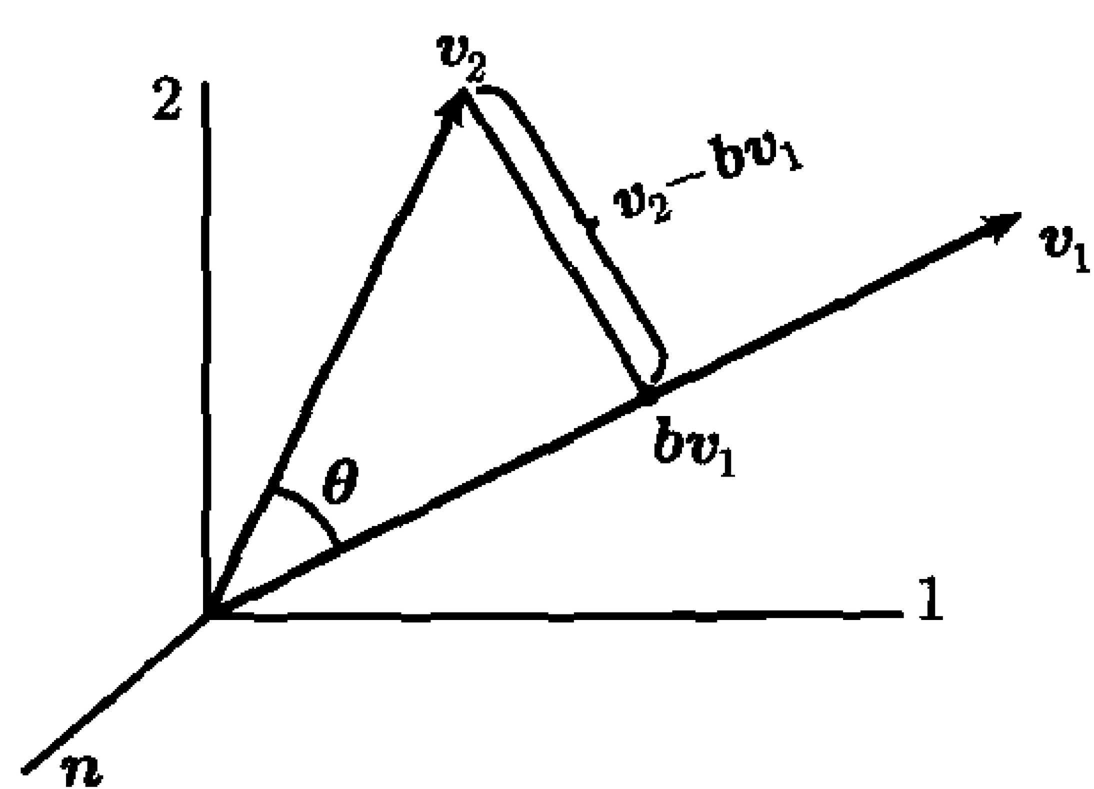


图 4.1

为了求 $\cos\theta$ 的分布, 我们可以先求出 $\cot\theta$ 的分布. 如 3.2 节证明的, 如果我们设 $b = \boldsymbol{v}_2'\boldsymbol{v}_1/\boldsymbol{v}_1'\boldsymbol{v}_1$, 则 $\boldsymbol{v}_2 - b\boldsymbol{v}_1$ 正交于 $\boldsymbol{v}_1$ 且

$$\cot\theta = \frac{b\|\boldsymbol{v}_1\|}{\|\boldsymbol{v}_2 - b\boldsymbol{v}_1\|}. \tag{8}$$

如果 $\boldsymbol{v}_1$ 固定, 我们可以旋转坐标轴使得第一个坐标轴沿 $\boldsymbol{v}_1$ 方向. 则 $b\boldsymbol{v}_1$ 就是第一个坐标和原点的差, 且 $\boldsymbol{v}_2 - b\boldsymbol{v}_1$ 的第一个这样的坐标等于零. 我们可以证明当 $\rho = 0$ 时 $\cot\theta$ 等比于一个 t 变量.

我们利用下面的引理.

引理 4.2.1 如果 $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_n$ 独立分布, 若 $\boldsymbol{Y}_\alpha = (\boldsymbol{Y}_\alpha^{(1)\prime}, \boldsymbol{Y}_\alpha^{(2)\prime})$ 有密度 $f(\boldsymbol{y}_\alpha)$, 且给定 $\boldsymbol{Y}_\alpha^{(1)} = \boldsymbol{y}_\alpha^{(1)}$ 时 $\boldsymbol{Y}_\alpha^{(2)}$ 的条件密度是 $f(\boldsymbol{y}_\alpha^{(2)}|\boldsymbol{y}_\alpha^{(1)})$, $\alpha = 1, \cdots, n$, 则在给定 $\boldsymbol{Y}_1^{(1)} = \boldsymbol{y}_1^{(1)}, \cdots, \boldsymbol{Y}_n^{(1)} = \boldsymbol{y}_n^{(1)}$ 时 $\boldsymbol{Y}_1^{(2)}, \cdots, \boldsymbol{Y}_n^{(2)}$ 的条件分布中, 随机向量 $\boldsymbol{Y}_1^{(2)}, \cdots, \boldsymbol{Y}_n^{(2)}$ 是独立的且 $\boldsymbol{Y}_\alpha^{(2)}$ 的密度是 $f(\boldsymbol{y}_\alpha^{(2)}|\boldsymbol{y}_\alpha^{(1)})$, $\alpha = 1, \cdots, n$.

证明 $\boldsymbol{Y}_1^{(1)}, \cdots, \boldsymbol{Y}_n^{(1)}$ 的边缘密度是 $\prod_{\alpha=1}^n f_1(\boldsymbol{y}_\alpha^{(1)})$, 其中 $f_1(\boldsymbol{y}_\alpha^{(1)})$ 是 $\boldsymbol{Y}_\alpha^{(1)}$ 的边缘密度, 且给定 $\boldsymbol{Y}_1^{(1)} = \boldsymbol{y}_1^{(1)}, \cdots, \boldsymbol{Y}_n^{(1)} = \boldsymbol{y}_n^{(1)}$ 时 $\boldsymbol{Y}_1^{(2)}, \cdots, \boldsymbol{Y}_n^{(2)}$ 的条件密度是

$$\frac{\prod_{\alpha=1}^n f(\boldsymbol{y}_\alpha)}{\prod_{\alpha=1}^n f_1(\boldsymbol{y}_\alpha^{(1)})} = \prod_{\alpha=1}^n \frac{f(\boldsymbol{y}_\alpha)}{f_1(\boldsymbol{y}_\alpha^{(1)})} = \prod_{\alpha=1}^n f(\boldsymbol{y}_\alpha^{(2)}|\boldsymbol{y}_\alpha^{(1)}). \qquad \blacksquare \tag{9}$$

用 $\boldsymbol{V}_i = (Z_{i1}, \cdots, Z_{in})'$, $i = 1, 2$, 表示随机向量. 给定 $Z_{1\alpha} = z_{1\alpha}$ 时 $Z_{2\alpha}$ 的条件分布是 $N(\beta z_{1\alpha}, \sigma^2)$, 其中 $\beta = \rho\sigma_2/\sigma_1$ 和 $\sigma^2 = \sigma_2^2(1-\rho^2)$. (见 2.5 节.) 因为 $Z_{2\alpha}$ 独立, 所以给定 $\boldsymbol{V}_1 = \boldsymbol{v}_1$ 时 $\boldsymbol{V}_2$ 的密度是 $N(\beta\boldsymbol{v}_1, \sigma^2\boldsymbol{I})$. 设 $b = \boldsymbol{V}_2'\boldsymbol{v}_i/\boldsymbol{v}_1'\boldsymbol{v}_1 (= a_{21}/a_{11})$, 所以 $b\boldsymbol{v}_1'(\boldsymbol{V}_2 - b\boldsymbol{v}_1) = 0$, 且设 $U = (\boldsymbol{V}_2 - b\boldsymbol{v}_1)'(\boldsymbol{V}_2 - b\boldsymbol{v}_1) = \boldsymbol{V}_2'\boldsymbol{V}_2 - b^2\boldsymbol{v}_1'\boldsymbol{v}_1 (= a_{22} - a_{12}^2/a_{11})$. 则 $\cot\theta = b\sqrt{a_{11}/U}$. 坐标轴的旋转需要选择一个第一行为 $(1/c)\boldsymbol{v}_1'$ 的 $n \times n$ 正交矩阵 $\boldsymbol{C}$, 其中 $c^2 = \boldsymbol{v}_1'\boldsymbol{v}_1$.

我们现在应用定理 3.3.1, 取 $\boldsymbol{X}_\alpha = Z_{2\alpha}$. 设 $Y_\alpha = \sum_\beta c_{\alpha\beta} Z_{2\beta}$, $\alpha = 1, \cdots, n$, 则 $Y_1, \cdots, Y_n$ 是独立正态分布的, 其方差为 σ^2, 均值为

$$\mathrm{E}(Y_1) = \sum_{\gamma=1}^n c_{1\gamma}\beta z_{1\gamma} = \frac{\beta}{c}\sum_{\gamma=1}^n z_{1\gamma}^2 = \beta c, \tag{10}$$

$$\mathrm{E}(Y_\alpha) = \sum_{\gamma=1}^n c_{\alpha\gamma}\beta z_{1\gamma} = \beta c\sum_{\gamma=1}^n c_{\alpha\gamma}c_{1\gamma} = 0, \quad \alpha \neq 1. \tag{11}$$

我们有 $b = \sum_{\alpha=1}^n Z_{2\alpha}z_{1\alpha}/\sum_{\alpha=1}^n z_{1\alpha}^2 = c\sum_{\alpha=1}^n Z_{2\alpha}c_{1\alpha}/c^2 = Y_1/c$, 并且从引理 3.3.1 知

$$U = \sum_{\alpha=1}^n Z_{2\alpha}^2 - b^2\sum_{\alpha=1}^n z_{1\alpha}^2 = \sum_{\alpha=1}^n Y_\alpha^2 - Y_1^2 = \sum_{\alpha=2}^n Y_\alpha^2 \tag{12}$$

与 b 独立. 则 U/σ^2 服从自由度为 $n-1$ 的 χ^2 分布.

引理 4.2.2 如果 $(Z_{1\alpha}, Z_{2\alpha}), \alpha=1,\cdots,n$ 独立, 每一对都有密度 (7), 则给定 $Z_{1\alpha}=z_{1\alpha}, \alpha=1,\cdots,n$ 时 $b=\sum_{\alpha=1}^n Z_{2\alpha}Z_{1\alpha}/\sum_{\alpha=1}^n Z_{1\alpha}^2$ 和 $U/\sigma^2=\sum_{\alpha=1}^n(Z_{2\alpha}-bZ_{1\alpha})^2/\sigma^2$ 的条件分布分别是 $N(\beta, \sigma^2/c^2)(c^2=\sum_{\alpha=1}^n z_{1\alpha}^2)$ 和自由度为 $n-1$ 的 χ^2 分布, 且 b 和 U 独立.

如果 $\rho=0$, 则 $\beta=0$ 且 b 条件地服从分布 $N(0, \sigma^2/c^2)$, 并且

$$\frac{cb/\sigma}{\sqrt{\frac{U/\sigma^2}{n-1}}}=\frac{cb}{\sqrt{\frac{U}{n-1}}} \tag{13}$$

服从自由度为 $n-1$ 的条件 t 分布. (见习题 4.27.) 但是, 这个随机变量是

$$\begin{aligned}\sqrt{n-1}\frac{\sqrt{a_{11}}a_{12}/a_{11}}{\sqrt{a_{22}-a_{12}^2/a_{11}}}&=\sqrt{n-1}\frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1-[a_{12}^2/(a_{11}a_{22})]}}\\&=\sqrt{n-1}\frac{r}{\sqrt{1-r^2}}.\end{aligned} \tag{14}$$

因此 $\sqrt{n-1}r/\sqrt{1-r^2}$ 服从自由度为 $n-1$ 的条件 t 分布. t 的密度是

$$\frac{\Gamma(\frac{1}{2}n)}{\sqrt{n-1}\Gamma[\frac{1}{2}(n-1)]\sqrt{\pi}}\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n}, \tag{15}$$

$W=r/\sqrt{1-r^2}$ 的密度是

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n-1)]\sqrt{\pi}}(1+w^2)^{-\frac{1}{2}n}. \tag{16}$$

因为 $w=r(1-r^2)^{-\frac{1}{2}}$, 所以我们有 $\mathrm{d}w/\mathrm{d}r=(1-r^2)^{-\frac{3}{2}}$. 因此, r 的密度是 (用 $N-1$ 代替 n)

$$\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-2)]\sqrt{\pi}}(1-r^2)^{\frac{1}{2}(N-4)}. \tag{17}$$

注意, (17) 是固定 $\boldsymbol{v}_1$ 时 r 的条件密度. 但是, 因为 (17) 不依赖 $\boldsymbol{v}_1$, 它也是 r 的边缘密度.

定理 4.2.1 设 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N$ 相互独立, 每一个都服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 如果 $\rho_{ij}=0$, 则由 (1) 定义的 r_{ij} 的密度是 (17).

从 (17) 我们看到这个密度是关于原点对称的. 当 $N>4$ 时, 在 $r=0$ 点有一个众数且它与 r 轴在 ±1 点的接触度当 N 为奇数时是 $\frac{1}{2}(N-5)$, 当 N 为偶数时是 $\frac{1}{2}N-3$. 因为此密度是偶函数, 从而奇数矩为零, 特别地, 均值是零. 偶数矩由积分求得 (令 $x=r^2$ 且利用贝塔函数的定义). 读者可以证明 $\mathrm{E}(r^{2m})=\Gamma[\frac{1}{2}(N-1)]\Gamma(m+\frac{1}{2})/\{\sqrt{\pi}\Gamma[\frac{1}{2}(N-1)+m]\}$, 特别地, 其方差为 $1/(N-1)$.

定理 4.2.1 最重要的用处是求关于一对变量不相关的假设检验的分位数. 对一些特殊的 (i,j) 考虑假设

$$H: \rho_{ij}=0. \tag{18}$$

如果相应的样本相关系数与零差别很大, 那么有理由拒绝这个假设. 现在我们怎样定义所谓的 "差别很大"?

假设我们感兴趣的是检验 H 及其备择假设 $\rho_{ij} > 0$. 如果样本相关系数 r_{ij} 大于某一个数 r_0, 那么我们拒绝 H. 当 H 为真时, 拒绝 H 的概率是

$$\int_{r_0}^{1} k_N(r)\mathrm{d}r, \tag{19}$$

其中 $k_N(r)$ 是 (17), 即基于 N 个观测的一个相关系数的密度. 我们选择 r_0 使得 (19) 是所期望的显著性水平. 如果我们检验 H 及其备择假设 $\rho_{ij} < 0$, 则当 $r_{ij} < -r_0$ 时, 拒绝 H.

现在假设我们关心的备择是 $\rho_{ij} \neq 0$; 即 ρ_{ij} 既可为正又可为负. 则当 $r_{ij} > r_1$ 或 $r_{ij} < -r_1$ 时, 我们拒绝假设 H. 当 H 为真时, 拒绝概率是

$$\int_{-1}^{-r_1} k_N(r)\mathrm{d}r + \int_{r_1}^{1} k_N(r)\mathrm{d}r. \tag{20}$$

选择 r_1 使得 (20) 为期望的显著性水平.

许多书里都给出了分位数 r_1 的选择, 包括 Fisher and Yates (1945) 中的表 VI, 表 VI 中的指标 n 等于我们的 $N-2$. 因为 $\sqrt{N-2}r/\sqrt{1-r^2}$ 服从自由度为 $N-2$ 的 t 分布, 所以同样可以使用 t 表. 对于备择 $\rho_{ij} \neq 0$, 如果

$$\sqrt{N-2}\frac{|r_{ij}|}{\sqrt{1-r_{ij}^2}} > t_{N-2}(\alpha), \tag{21}$$

其中 $t_{N-2}(\alpha)$ 是自由度为 $N-2$ 的 t 统计量在显著性水平为 α 时的双尾分位数, 那么拒绝 H. 对于备择 $\rho_{ij} > 0$, 如果

$$\sqrt{N-2}\frac{r_{ij}}{\sqrt{1-r_{ij}^2}} > t_{N-2}(2\alpha), \tag{22}$$

则拒绝 H.

从 (13) 和 (14) 我们看到 $\sqrt{N-2}r/\sqrt{1-r^2}$ 对于检验 $\boldsymbol{V}_2$ 在 $\boldsymbol{v}_1$ 上的回归是零的假设是一个合适的统计量. 从原始观测 $\{x_{i\alpha}\}$ 的角度看, 我们有

$$\sqrt{N-2}\frac{r}{\sqrt{1-r^2}} = \frac{b\sqrt{\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2}}{\sqrt{\sum_{\alpha=1}^{N}[x_{2\alpha}-\bar{x}_2-b(x_{1\alpha}-\bar{x}_1)]^2/(N-2)}}, \tag{23}$$

其中 $b=\sum_{\alpha=1}^{N}(x_{2\alpha}-\bar{x}_2)(x_{1\alpha}-\bar{x}_1)/\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2$ 是 $x_{2\alpha}$ 在 $x_{1\alpha}$ 上的最小二乘回归系数. 可以看出, 检验 $\rho_{12}=0$ 等价于检验 $\boldsymbol{X}_2$ 在 x_1 上的回归是零 (即 $\rho_{12}\sigma_2/\sigma_1=0$).

为了说明这个方法, 我们考虑 3.2 节给出的一个例子. 我们来检验两种药物功效不相关的原假设及它们正相关的备择假设. 我们可以选择 5% 的显著性水平. 对 $N=10$, 5% 分位数 (r_0) 是 0.5494. 我们观测到的相关系数 0.7952 是显著的, 所以我们拒绝两种药物功效独立的假设.

4.2.2 当总体相关系数非零时的分布，假设检验和置信区间

为了求总体相关系数非零时样本相关系数的分布，我们需要先得到 a_{11}，a_{12} 和 a_{22} 的联合密度. 在 4.2.1 节，我们看到，固定 $\boldsymbol{v}_1$ 的条件下，随机变量 $b=a_{12}/a_{11}$ 和 $U/\sigma^2=(a_{22}-a_{12}^2/a_{11})/\sigma^2$ 分别独立地服从分布 $N(\beta,\sigma^2/c^2)$ 和自由度为 $n-1$ 的 χ^2 分布. 用 $g_{n-1}(u)$ 表示 χ^2 分布的密度，我们记 b 和 U 的条件密度为 $n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)/\sigma^2$. $\boldsymbol{V}_1$, b 和 U 的联合密度是 $n(\boldsymbol{v}_1|\boldsymbol{0},\sigma_1^2\boldsymbol{I})n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)/\sigma^2$. $\boldsymbol{V}_1'\boldsymbol{V}_1/\sigma_1^2=a_{11}/\sigma_1^2$ 的边缘密度是 $g_n(u)$，即 a_{11} 的密度是

$$\frac{1}{\sigma_1^2}g_n\left(\frac{a_{11}}{\sigma_1^2}\right)=\int\cdots\int_{\boldsymbol{v}_1'\boldsymbol{v}_1=a_{11}} n(\boldsymbol{v}_1|\boldsymbol{0},\sigma_1^2\boldsymbol{I})\mathrm{d}W. \tag{24}$$

其中 $\mathrm{d}W$ 是一个合适的体积元.

积分是在球面 $\boldsymbol{v}_1'\boldsymbol{v}_1=a_{11}$ 上的，因此，$\mathrm{d}W$ 是此球面上的一个面积元. (在定义 $\mathrm{d}W$ 时用角坐标，见习题 7.1.) 因此 b, U 和 a_{11} 的联合密度是

$$\begin{aligned}
&\int\cdots\int_{\boldsymbol{v}_1'\boldsymbol{v}_1=a_{11}} n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)\frac{1}{\sigma^2}n(\boldsymbol{v}_1|\boldsymbol{0},\sigma_1^2\boldsymbol{I})\mathrm{d}W \\
&=\frac{g_n(a_{11}/\sigma_1^2)n(b|\beta,\sigma^2/a_{11})g_{n-1}(u/\sigma^2)}{\sigma_1^2\sigma^2} \\
&=\frac{(a_{11})^{\frac{1}{2}n-1}}{(2\sigma_1^2)^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\exp\left(-\frac{a_{11}}{2\sigma_1^2}\right)\frac{\sqrt{a_{11}}}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{a_{11}}{2\sigma^2}(b-\beta)^2\right] \\
&\quad\cdot\frac{1}{(2\sigma^2)^{\frac{1}{2}(n-1)}\Gamma[\frac{1}{2}(n-1)]}u^{\frac{1}{2}(n-3)}\exp\left(-\frac{u}{2\sigma^2}\right).
\end{aligned} \tag{25}$$

现在令 $b=a_{12}/a_{11}$, $U=a_{22}-a_{12}^2/a_{11}$. 雅可比行列式是

$$\left|\frac{\partial(b,u)}{\partial(a_{12},a_{22})}\right|=\begin{vmatrix}\frac{1}{a_{11}} & 0\\ -2\frac{a_{12}}{a_{11}} & 1\end{vmatrix}=\frac{1}{a_{11}}. \tag{26}$$

因此，当 $a_{11}\geqslant 0$, $a_{22}\geqslant 0$ 和 $a_{11}a_{22}-a_{12}^2\geqslant 0$ 时，a_{11}, a_{12} 和 a_{22} 的密度是

$$\frac{a_{11}^{\frac{1}{2}(n-3)}\left(\frac{a_{11}a_{22}-a_{12}^2}{a_{11}}\right)^{\frac{1}{2}(n-3)}\mathrm{e}^{-\frac{1}{2}Q}}{2^n\sigma_1^n\sigma_2^n(1-\rho^2)^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}, \tag{27}$$

其中

$$\begin{aligned}
Q&=\frac{a_{11}}{\sigma_1^2}+\frac{a_{11}}{\sigma^2}\left(\frac{a_{12}^2}{a_{11}^2}-2\rho\frac{\sigma_1\sigma_2}{\sigma_1^2}\frac{a_{12}}{a_{11}}+\frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_1^4}\right)+\frac{1}{\sigma^2}\left(a_{22}-\frac{a_{12}^2}{a_{11}}\right) \\
&=a_{11}\left[\frac{1}{\sigma_1^2}+\frac{\rho^2\sigma_1^2\sigma_2^2}{\sigma_1^4\sigma_2^2(1-\rho^2)}\right]-2a_{12}\frac{\rho\sigma_2}{\sigma_1\sigma_2^2(1-\rho^2)}+\frac{a_{22}}{\sigma_2^2(1-\rho^2)} \\
&=\frac{1}{1-\rho^2}\left(\frac{a_{11}}{\sigma_1^2}-2\rho\frac{a_{12}}{\sigma_1\sigma_2}+\frac{a_{22}}{\sigma_2^2}\right).
\end{aligned} \tag{28}$$

当 $\boldsymbol{A}$ 正定时, 这个密度可以记为

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(n-3)}\mathrm{e}^{-\frac{1}{2}Q}}{2^n|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}, \tag{29}$$

否则, 为 0. 这是第 7 章中 Wishart 密度的一个特殊情况.

我们想求

$$r=\frac{a_{12}}{\sqrt{a_{11}a_{22}}}=\frac{a_{12}/(\sigma_1\sigma_2)}{\sqrt{(a_{11}/\sigma_1^2)(a_{22}/\sigma_2^2)}}=\frac{a_{12}^*}{\sqrt{a_{11}^*a_{22}^*}}, \tag{30}$$

其中 $a_{11}^*=a_{11}/\sigma_1^2$, $a_{22}^*=a_{22}/\sigma_2^2$ 和 $a_{12}^*=a_{12}/(\sigma_1\sigma_2)$. 这个变换等价于令 $\sigma_1=\sigma_2=1$. 则 a_{11}, a_{22} 和 $r=a_{12}/\sqrt{a_{11}a_{22}}(\mathrm{d}a_{12}=\mathrm{d}r\sqrt{a_{11}a_{22}})$ 的密度是

$$\frac{a_{11}^{\frac{1}{2}n-1}a_{22}^{\frac{1}{2}n-1}(1-r^2)^{\frac{1}{2}(n-3)}\mathrm{e}^{-\frac{1}{2}Q}}{2^n(1-\rho^2)^{\frac{1}{2}n}\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}, \tag{31}$$

其中

$$Q=\frac{a_{11}-2\rho r\sqrt{a_{11}}\sqrt{a_{22}}+a_{22}}{1-\rho^2}. \tag{32}$$

为了求 r 的密度, 我们必须将 (31) 关于 a_{11} 和 a_{22} 在区间 0 到 ∞ 上积分. 不同的积分方法可使这个密度有不同的表达式. 我们这里用的方法是直观的. 我们展开指数部分:

$$\exp\left[\frac{\rho r\sqrt{a_{11}}\sqrt{a_{22}}}{(1-\rho^2)}\right]=\sum_{\alpha=0}^{\infty}\frac{(\rho r\sqrt{a_{11}}\sqrt{a_{22}})^\alpha}{\alpha!(1-\rho^2)^\alpha}. \tag{33}$$

则密度 (31) 是

$$\begin{aligned}&\frac{(1-r^2)^{\frac{1}{2}(n-3)}}{(1-\rho^2)^{\frac{1}{2}n}2^n\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\sum_{\alpha=0}^{\infty}\frac{(\rho r)^\alpha}{\alpha!(1-\rho^2)^\alpha}\\&\cdot\left\{\exp\left[-\frac{a_{11}}{2(1-\rho^2)}\right]a_{11}^{(n+\alpha)/2-1}\right\}\left\{\exp\left[-\frac{a_{22}}{2(1-\rho^2)}\right]a_{22}^{(n+\alpha)/2-1}\right\}.\end{aligned} \tag{34}$$

因为

$$\int_0^\infty a^{\frac{1}{2}(n+\alpha)-1}\exp\left[-\frac{a}{2(1-\rho^2)}\right]\mathrm{d}a=\Gamma[\frac{1}{2}(n+\alpha)][2(1-\rho^2)]^{\frac{1}{2}(n+\alpha)}, \tag{35}$$

所以 (34) 的积分 (允许逐项积分) 是

$$\begin{aligned}&\frac{(1-r^2)^{\frac{1}{2}(n-3)}}{(1-\rho^2)^{\frac{1}{2}n}2^n\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\\&\cdot\sum_{\alpha=0}^{\infty}\frac{(\rho r)^\alpha}{\alpha!(1-\rho^2)^\alpha}\Gamma^2\left[\frac{1}{2}(n+\alpha)\right]2^{n+\alpha}(1-\rho^2)^{n+\alpha}\\=&\frac{(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)\Gamma[\frac{1}{2}(n-1)]}\sum_{\alpha=0}^{\infty}\frac{(2\rho r)^\alpha}{\alpha!}\Gamma^2\left[\frac{1}{2}(n+\alpha)\right].\end{aligned} \tag{36}$$

伽玛函数的倍量公式是

$$\Gamma(2z)=\frac{2^{2z-1}\Gamma(z)(z+\frac{1}{2})}{\sqrt{\pi}}. \tag{37}$$

可以用它修正 (36) 中的常数.

定理 4.2.2　对于来自于相关系数为 ρ 的二元正态分布的容量为 N 的样本, 其相关系数的分布密度为

$$\frac{2^{n-2}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{(n-2)!\pi}\sum_{\alpha=0}^{\infty}\frac{(2\rho r)^\alpha}{\alpha!}\Gamma^2\left[\frac{1}{2}(n+\alpha)\right],\quad -1\leqslant r\leqslant 1, \tag{38}$$

其中 $n=N-1$.

Fisher (1915) 第一次求出 r 的分布. 他也给出了该密度的另一个形式,

$$\frac{(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}}{\pi(n-2)!}\left[\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}\left\{\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}\right\}\bigg|_{x=r\rho}\right]. \tag{39}$$

见习题 4.24.

Hotelling (1953) 对 r 的分布做了详细的研究. 他推荐下面的形式:

$$\frac{n-1}{\sqrt{2\pi}}\frac{\Gamma(n)}{\Gamma(n+\frac{1}{2})}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)} \cdot(1-\rho r)^{-n+\frac{1}{2}}F\left(\frac{1}{2},\frac{1}{2};n+\frac{1}{2};\frac{1+\rho r}{2}\right), \tag{40}$$

其中

$$F(a,b;c;x)=\sum_{j=0}^{\infty}\frac{\Gamma(a+j)}{\Gamma(a)}\frac{\Gamma(b+j)}{\Gamma(b)}\frac{\Gamma(c)}{\Gamma(c+j)}\frac{x^j}{j!} \tag{41}$$

是一个超几何函数. (见习题 4.25.) 序列 (40) 的收敛速度比序列 (38) 快. Hotelling 讨论了积分密度的方法并计算了 r 的矩.

David (1938) 用表列出了[①] $\rho=0(.1).9, N=3(1)25,50,100,200,400$ 和 $r^*=-1(.05)1$(David 的 n 是我们的 N) 时, r 的累积分布

$$\Pr\{r\leqslant r^*\}=F(r^*|N,\rho). \tag{42}$$

从密度 (38) 可以显然得到 $F(r^*|N,\rho)=1-F(-r^*|N,-\rho)$, 因为 r,ρ 的密度等于 $-r,-\rho$ 的密度. 这些表可以用在许多统计方法中.

首先, 我们考虑用样本检验假设

$$H:\rho=\rho_0 \tag{43}$$

的问题. 如果备择是 $\rho>\rho_0$, 当样本相关系数大于 r_0 时我们拒绝原假设, 其中 r_0 满足 $1-F(r_0|N,\rho_0)=\alpha$, α 是显著性水平. 如果备择是 $\rho<\rho_0$, 当样本相关系数小于 r_0' 时我们拒绝原假设, 其中 r_0' 满足 $F(r_0'|N,\rho_0)=\alpha$. 如果备择是 $\rho\neq\rho_0$, 则拒

① $\rho=0(.1).9$ 表示 $\rho=0.1,0.2,\cdots,0.9$.

绝区域是 $r > r_1$ 和 $r < r_1'$, 其中 r_1 和 r_1' 满足 $[1 - F(r_1|N, \rho_0)] + F(r_1'|N, \rho_0) = \alpha$. David 建议选择满足 $[1 - F(r_1|N, \rho_0)] = F(r_1'|N, \rho_0) = \frac{1}{2}\alpha$ 的 r_1 和 r_1'. 她证明了 (1937) 当 $N \geqslant 10, |\rho| \leqslant 0.8$ 时这个判别区域近似为 H 的无偏检验的区域, 即一个功效函数在 ρ_0 点有极小值的检验.

需要指出的是任意基于 r 的检验在位移变换和尺度变换下是不变的, 即 $x_{i\alpha}^* = b_i x_{i\alpha} + c_i$, $b_i > 0$, $i = 1, 2$, $\alpha = 1, \cdots, N$, 且 r 本质上是充分统计量唯一的不变量 (习题 3.7). 上面检验 $H: \rho = \rho_0$ 及其备择 $\rho > \rho_0$ 的方法是所有不变检验中一致最优的. (见习题 4.16~4.18.)

例如, 用一组容量为 15 的样本的相关系数以显著性水平 5% 检验假设 $\rho = 0.5$ 及备择 $\rho \neq 0.5$. 我们在 David 的表中找到 (用插值方法) $F(0.027|15, 0.5) = 0.025$ 和 $F(0.805|15, 0.5) = 0.975$. 因此, 当 r 小于 0.027 或大于 0.805 时我们拒绝原假设.

接下来, 我们用 David 的表计算一个相关性检验的功效函数. 如果 H 的拒绝区域是 $r > r_1$ 和 $r < r_1'$, 则检验的功效函数是真正的相关系数 ρ 的函数, 即 $[1 - F(r_1|N, \rho)] + [F(r_1'|N, \rho)]$, 这是当总体相关系数为 ρ 时拒绝原假设的概率.

例如, 考虑求上节中 $\rho = 0$ 的检验的功效函数. 一个显著性水平为 5% 的拒绝区域 (单边) 是 $r \geqslant 0.5494$. 拒绝的概率在表 4.1 中给出. 图 4.2 是功效函数的图.

表 4.1　功效函数

ρ	概率	ρ	概率
−1.0	0.0000	0.2	0.1376
−0.8	0.0000	0.4	0.3215
−0.6	0.0004	0.6	0.6235
−0.4	0.0032	0.8	0.9279
−0.2	0.0417	1.0	1.0000
0.0	0.0500		

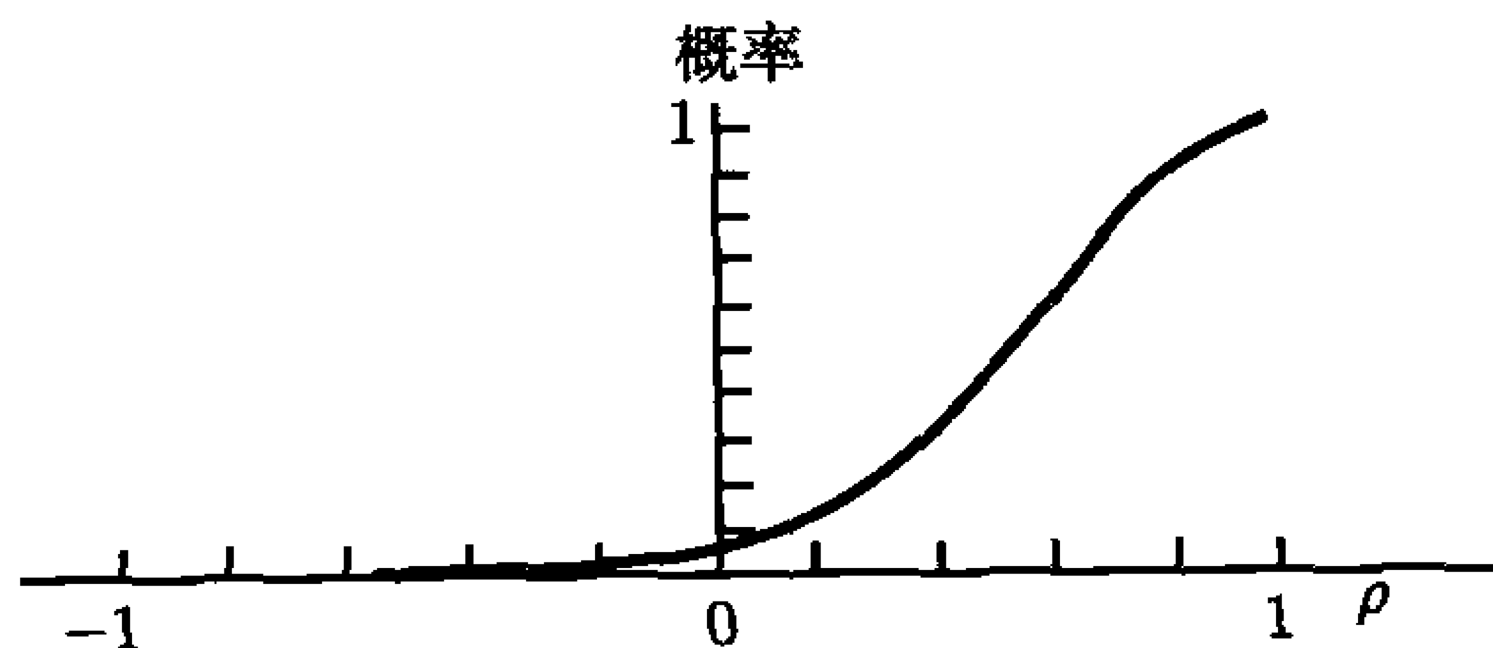


图 4.2　一个功效函数

然后, David 计算了 ρ 的置信区间. 对于给定的 N, r_1' (定义一个分位数) 是 ρ 的函数, 记做 $f_1(\rho)$, 且 r_1 是 ρ 的另一个函数, 记做 $f_2(\rho)$, 满足

$$\Pr\{f_1(\rho) < r < f_2(\rho)|\rho\} = 1 - \alpha. \tag{44}$$

显然, 如果 r_1 和 r_1' 满足 $1 - F(r_1|N, \rho) = \frac{1}{2}\alpha = F(r_1'|N, \rho)$, 那么 $f_1(\rho)$ 和 $f_2(\rho)$ 是

ρ 的单调递增函数. 如果 $\rho = f_i^{-1}(r)$ 是 $r = f_i(\rho)(i=1,2)$ 的逆, 则不等式 $f_1(\rho) < r$ 等价于[①] $\rho < f_1^{-1}(r)$ 且 $r < f_2(\rho)$ 等价于 $f_2^{-1} < \rho$. 因此 (44) 可以记为

$$\Pr\{f_2^{-1}(r) < \rho < f_1^{-1}(r)|\rho\} = 1 - \alpha. \tag{45}$$

这个等式说明我们抽取到使得区间 $(f_2^{-1}(r), f_1^{-1}(r))$ 覆盖参数 ρ 的样本的概率是 $1-\alpha$. 因此, 这个区间是 ρ 的一个置信系数为 $1-\alpha$ 的置信区间. 给定 N 和 α, 曲线 $r = f_1(\rho)$ 和 $r = f_2(\rho)$ 如图 4.3 所示. 在检验 $\rho = \rho_0$ 中, 直线 $\rho = \rho_0$ 和两条曲线的交点给出了分位数 r_1 和 r_1'. 在基于样本相关系数 r^* 建立 ρ 的置信区域中, 我们由直线 $r = r^*$ 和两曲线的交点求得限 $f_2^{-1}(r^*)$ 和 $f_1^{-1}(r^*)$. David 给出了这些曲线在 $\alpha = 0.1, 0.05, 0.02$ 和 0.01, N 为不同值时的情况.

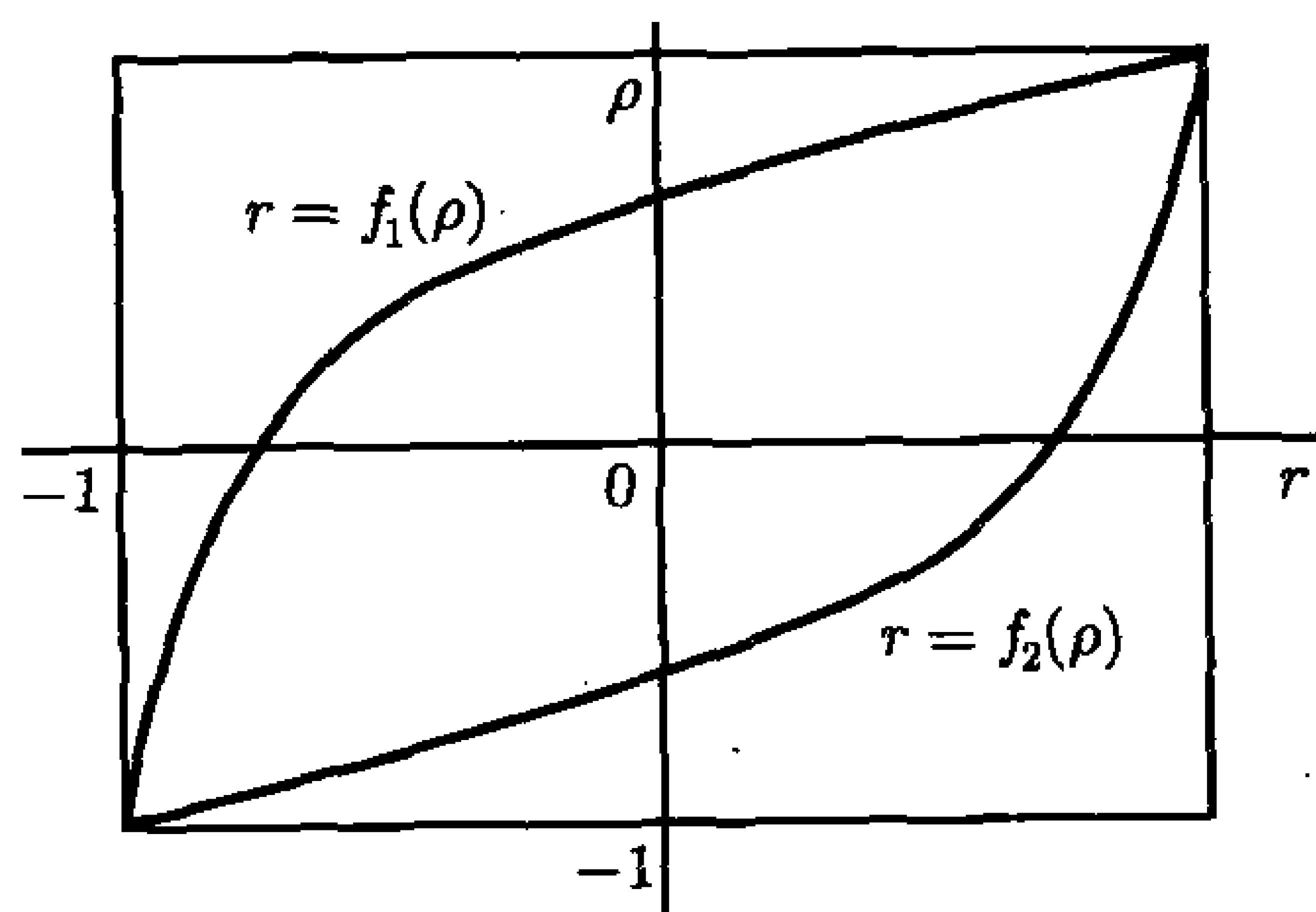


图 4.3

在求置信区间时, $F(r|N,\rho)$ 的表也可以用来替代这些曲线. 给定样本值 r^*, $f_1^{-1}(r^*)$ 是满足 $\frac{1}{2}\alpha = \Pr\{r \leqslant r^*|\rho\} = F(r^*|N,\rho)$ 的 ρ 值. 相似地, $f_2^{-1}(r^*)$ 是满足 $\frac{1}{2}\alpha = \Pr\{r \geqslant r^*|\rho\} = 1 - F(r^*|N,\rho)$ 的 ρ 值. 这两个 ρ 值之间的区间, $(f_2^{-1}(r^*), f_1^{-1}(r^*))$, 就是所求的置信区间.

例如, 对于基于样本相关系数 0.7952 的容量为 10 的样本, 寻求置信系数为 0.95 的置信区间. 用 David 的图 II, 我们求出两个限是 0.34 和 0.94. 于是我们得到置信度为 95% 的置信区间 $0.34 < \rho < 0.94$.

定义 4.2.1 *设 $L(\boldsymbol{x}, \boldsymbol{\theta})$ 是观测向量 $\boldsymbol{x}$ 和参数向量 $\boldsymbol{\theta} \in \Omega$ 的似然函数. 设原假设由 Ω 的真子集 ω 定义. 似然比准则是*

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta}\in\omega} L(\boldsymbol{x}, \boldsymbol{\theta})}{\sup_{\boldsymbol{\theta}\in\Omega} L(\boldsymbol{x}, \boldsymbol{\theta})}. \tag{46}$$

似然比检验是当 $\lambda(\boldsymbol{x})$ 小于预先给定的常数时拒绝原假设的方法.

直观地, 当观测在原假设中最佳参数选择下的密度远远小于其在最佳无约束参数选择下的密度时, 就拒绝原假设. 似然比检验有许多所期望的特征, 例如, 见

① 第一条曲线上的点 $(f_1(\rho), \rho)$ 在 (r, ρ) 的左边而点 $(r, f_1^{-1}(r))$ 在 (r, ρ) 的上面.

Lehmann (1959). Wald (1943) 证明了一些不错的渐近性质. 对于大多数关于多元正态分布的检验, 似然比检验是合适的, 常常也是最优的.

我们来考虑基于来自二元正态分布的样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的检验 $\rho=\rho_0$. 集合 Ω 由满足 $\sigma_1>0,\sigma_2>0$ 和 $-1<\rho<1$ 的 $\mu_1,\mu_2,\sigma_1,\sigma_2$ 和 ρ 组成. 集合 ω 是满足 $\rho=\rho_0$ 的子集. Ω 中极大化的似然是 (由引理 3.2.2 和引理 3.2.3)

$$\max_{\Omega} L=\frac{N^N \mathrm{e}^{-N}}{(2\pi)^N(1-r^2)^{\frac{1}{2}N}a_{11}^{N/2}a_{22}^{N/2}}. \tag{47}$$

在原假设下的似然函数是

$$\frac{1}{(2\pi)^N(1-\rho_0^2)^{\frac{1}{2}N}(\sigma^2)^N}\exp\left[-\frac{a_{11}/\tau+\tau a_{22}-2\rho_0 a_{12}}{2\sigma^2(1-\rho_0^2)}\right], \tag{48}$$

其中 $\sigma^2=\sigma_1\sigma_2$ 和 $\tau=\sigma_1/\sigma_2$. (48) 关于 τ 的极大值在点 $\hat{\tau}=\sqrt{a_{11}}/\sqrt{a_{22}}$ 取到. 集中似然是

$$\frac{1}{(2\pi)^N(1-\rho_0^2)^{\frac{1}{2}N}(\sigma^2)^N}\exp\left[-\frac{\sqrt{a_{11}}\sqrt{a_{22}}(1-\rho_0 r)}{\sigma^2(1-\rho_0^2)}\right], \tag{49}$$

(49) 的极大值出现在

$$\hat{\sigma}^2=\frac{a_{11}^{\frac{1}{2}}a_{22}^{\frac{1}{2}}(1-\rho_0 r)}{N(1-\rho_0^2)}. \tag{50}$$

因此, 似然比准则是

$$\frac{\max_{\omega} L}{\max_{\Omega} L}=\frac{(1-\rho_0^2)^{\frac{1}{2}N}(1-r^2)^{\frac{1}{2}N}}{(1-\rho_0 r)^N}=\left[\frac{(1-\rho_0^2)(1-r^2)}{(1-\rho_0 r)^2}\right]^{\frac{1}{2}N}. \tag{51}$$

这个似然比检验是 $(1-\rho_0^2)(1-r^2)(1-\rho_0 r)^{-2}<c$, 其中 c 使得当样本取自相关系数为 ρ_0 的正态总体时, 不等式的概率是给定的显著性水平. 判别区域可以等价地记为

$$(\rho_0^2 c-\rho_0^2+1)r^2-2\rho_0 cr+c-1+\rho_0^2>0, \tag{52}$$

或者

$$r>\frac{\rho_0 c+(1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c+1-\rho_0^2},\quad r<\frac{\rho_0 c-(1-\rho_0^2)\sqrt{1-c}}{\rho_0^2 c+1-\rho_0^2}. \tag{53}$$

因此, $H:\rho=\rho_0$ 及备择 $\rho\neq\rho_0$ 的似然比检验有一个形如 $r>r_1$ 和 $r<r_1'$ 的拒绝区域. 但是 r_1 和 r_1' 不满足当 H 为真时, 每一个不等式的概率都为 $\alpha/2$, 而是由 (53) 而来, 其中 c 使得这两个不等式的概率和为 α.

4.2.3 样本相关系数的渐近分布和 Fisher z

本节我们将证明, 随着样本容量的增加, 样本相关系数的分布趋于正态分布. 样本相关系数的特殊函数 (Fisher z) 的分布, 有着和总体相关系数近似独立的方差 [Fisher (1921)], 但是它更快地趋于正态.

我们对样本相关系数

$$r(n) = \frac{A_{ij}(n)}{\sqrt{A_{ii}(n)A_{jj}(n)}} \tag{54}$$

特别感兴趣, 其中 $i \neq j$. 它可以记为

$$r(n) = \frac{C_{ij}(n)}{\sqrt{C_{ii}(n)C_{jj}(n)}}, \tag{55}$$

其中 $C_{gh}(n) = A_{gh}(n)/\sqrt{\sigma_{gg}\sigma_{hh}}$. $C_{ii}(n)$, $C_{jj}(n)$, $C_{ij}(n)$ 如同下面矩阵中不同元素的分布:

$$\sum_{\alpha=1}^{n} \begin{pmatrix} Z_{i\alpha}^* \\ Z_{j\alpha}^* \end{pmatrix} (Z_{i\alpha}^*, Z_{j\alpha}^*) = \sum_{\alpha=1}^{n} \begin{pmatrix} Z_{i\alpha}/\sqrt{\sigma_{ii}} \\ Z_{j\alpha}/\sqrt{\sigma_{jj}} \end{pmatrix} (Z_{i\alpha}/\sqrt{\sigma_{ii}}, Z_{j\alpha}/\sqrt{\sigma_{jj}}), \tag{56}$$

其中 $(Z_{i\alpha}^*, Z_{j\alpha}^*)$ 是独立的, 都服从分布

$$N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

其中

$$\rho = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}.$$

设

$$\boldsymbol{U}(n) = \frac{1}{n}\begin{pmatrix} C_{ii}(n) \\ C_{jj}(n) \\ C_{ij}(n) \end{pmatrix}, \tag{57}$$

$$\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ \rho \end{pmatrix}. \tag{58}$$

则由定理 3.4.4, 向量 $\sqrt{n}[\boldsymbol{U}(n) - \boldsymbol{b}]$ 有极限正态分布, 其均值为 $\mathbf{0}$, 协方差阵为

$$\begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix}. \tag{59}$$

现在我们需要一般定理.

定理 4.2.3 设 $\{\boldsymbol{U}(n)\}$ 是 m 维随机向量序列且 $\boldsymbol{b}$ 是一个固定向量, 满足当 $n \to \infty$ 时 $\sqrt{n}[\boldsymbol{U}(n) - \boldsymbol{b}]$ 有极限分布 $N(\mathbf{0}, \mathbf{T})$. 设 $\boldsymbol{f}(\boldsymbol{u})$ 是 $\boldsymbol{u}$ 的一个向量值函数, 其中每一个分量 $f_j(\boldsymbol{u})$ 在 $\boldsymbol{u} = \boldsymbol{b}$ 有非零微分, 则令 $\partial f_j(\boldsymbol{u})/\partial u_i|_{\boldsymbol{u}=\boldsymbol{b}}$ 为 $\boldsymbol{\Phi_b}$ 的第 i, j 个元素. 则 $\sqrt{n}\{\boldsymbol{f}[\boldsymbol{u}(n)] - \boldsymbol{f}(\boldsymbol{b})\}$ 有极限分布 $N(\mathbf{0}, \boldsymbol{\Phi}_{\boldsymbol{b}}'\mathbf{T}\boldsymbol{\Phi_b})$.

证明 见 Serfling (1980) 的 3.3 节, 或 Rao (1973) 的 6a.2 节. 一个函数 $g(\boldsymbol{u})$ 称为在 $\boldsymbol{b}$ 处有微分或完全可微, 如果在 $\boldsymbol{u} = \boldsymbol{b}$ 存在偏导数 $\partial g(\boldsymbol{u})/\partial u_i$ 且对每一个 $\varepsilon > 0$, 存在一个临域 $N_\varepsilon(\boldsymbol{b})$ 满足

$$\left|g(\boldsymbol{u})-g(\boldsymbol{b})-\sum_{i=1}^{m}\frac{\partial g(\boldsymbol{u})}{\partial u_i}(u_i-b_i)\right|\leqslant\varepsilon\|\boldsymbol{u}-\boldsymbol{b}\|\quad \text{对所有}\boldsymbol{u}\in N_{\varepsilon}(\boldsymbol{b}). \qquad \blacksquare\ (60)$$

显然 (57) 定义的 $\boldsymbol{U}(n)$ 满足定理的条件, 其中 $\boldsymbol{b}$ 和 $\boldsymbol{T}$ 分别由 (58) 和 (59) 定义, 函数

$$r=\frac{u_3}{\sqrt{u_1u_2}}=u_3u_1^{-\frac{1}{2}}u_2^{-\frac{1}{2}} \tag{61}$$

满足条件, $\boldsymbol{\Phi_b}$ 的元素是

$$\begin{aligned}
\left.\frac{\partial r}{\partial u_1}\right|_{\boldsymbol{u}=\boldsymbol{b}}&=-\frac{1}{2}u_3u_1^{-\frac{3}{2}}u_2^{-\frac{1}{2}}|_{\boldsymbol{u}=\boldsymbol{b}}=-\frac{1}{2}\rho,\\
\left.\frac{\partial r}{\partial u_2}\right|_{\boldsymbol{u}=\boldsymbol{b}}&=-\frac{1}{2}u_3u_1^{-\frac{3}{2}}u_2^{-\frac{1}{2}}|_{\boldsymbol{u}=\boldsymbol{b}}=-\frac{1}{2}\rho,\\
\left.\frac{\partial r}{\partial u_3}\right|_{\boldsymbol{u}=\boldsymbol{b}}&=u_1^{-\frac{1}{2}}u_2^{-\frac{1}{2}}|_{\boldsymbol{u}=\boldsymbol{b}}=1,
\end{aligned} \tag{62}$$

$f(\boldsymbol{b})=\rho$. $\sqrt{n}[r(n)-\rho]$ 的极限分布的方差是

$$\begin{aligned}
&\left(-\frac{1}{2}\rho,-\frac{1}{2}\rho,1\right)\begin{pmatrix}2 & 2\rho^2 & 2\rho\\ 2\rho^2 & 2 & 2\rho\\ 2\rho & 2\rho & 1+\rho^2\end{pmatrix}\begin{pmatrix}-\frac{1}{2}\rho\\ -\frac{1}{2}\rho\\ 1\end{pmatrix} \qquad (63)\\
=&(\rho-\rho^3,\rho-\rho^3,1-\rho^2)\begin{pmatrix}-\frac{1}{2}\rho\\ -\frac{1}{2}\rho\\ 1\end{pmatrix}\\
=&1-2\rho^2+\rho^4\\
=&(1-\rho^2)^2.
\end{aligned}$$

因此, 我们得到下面的定理.

定理 4.2.4　考虑来自相关系数为 ρ 的正态分布的样本, 其容量为 $N\ (=n+1)$. 如果 $r(n)$ 是其样本相关系数, 则 $\sqrt{n}[r(n)-\rho]/(1-\rho^2)$ [或 $\sqrt{N}[r(n)-\rho]/(1-\rho^2)$] 有极限分布 $N(0,1)$.

由定理 4.2.3 显知, 如果 $f(x)$ 在 $x=\rho$ 可微, 则 $\sqrt{n}[f(r)-f(\rho)]$ 有渐近正态分布, 其均值为零, 方差为

$$\left(\left.\frac{\partial f}{\partial x}\right|_{x=\rho}\right)^2(1-\rho^2)^2.$$

考虑一个有用函数, 它的渐近方差是独立于 ρ 的常数 (这里为单位 1). 这个函数满足等式

$$f'(\rho)=\frac{1}{1-\rho^2}=\frac{1}{2}\left(\frac{1}{1+\rho}+\frac{1}{1-\rho}\right). \tag{64}$$

因此 $f(\rho)$ 可取为 $\frac{1}{2}[\ln(1+\rho)-\ln(1-\rho)]=\frac{1}{2}[\ln(1+\rho)/(1-\rho)]$. 所谓的 Fisher z 是

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} = \tanh^{-1} r, \tag{65}$$

其中 $r = \tanh z = (e^z - e^{-z})/(e^z + e^{-z})$. 设

$$\zeta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}. \tag{66}$$

定理 4.2.5 *设 z 的定义为 (65), 其中 r 是容量为 $N(=n+1)$ 的样本的相关系数, 该样本来自相关系数为 ρ 的二元正态分布; 设 ζ 的定义为 (66). 则 $\sqrt{n}(z-\zeta)$ 有均值为 0, 方差为 1 的极限正态分布.*

可以证明一个更接近的近似

$$E(z) \sim \zeta + \frac{\rho}{2n}. \tag{67}$$

$$E(z-\zeta)^2 \sim \frac{1}{n-2} \sim E\left(z - \zeta - \frac{\rho}{2n}\right)^2. \tag{68}$$

后者从

$$E(z-\zeta)^2 = \frac{1}{n} + \frac{8-\rho^2}{4n^2} + \cdots \tag{69}$$

而来且当 ρ^2/n^2 很小时表现较好. Hotelling (1953) 给出了 z 的矩, 至阶数 n^{-3}. Fisher z 的一个重要性质是它接近正态的速度远大于 r. David (1938) 把列表中的概率和通过假定 z 为正态分布而计算出的概率进行对比. 她推荐 $N > 25$ 时, 视 z 服从均值和方差分别为 (67) 和 (68) 的正态分布. Konishi (1978a, 1978b, 1979) 也研究了 z. [Ruben (1966) 提出了另一个更复杂但更精确的方法.]

我们现在来说明怎样利用定理 4.2.5.

a. 假设我们希望基于容量为 N 的样本检验假设 $\rho = \rho_0$ 及其备择 $\rho \neq \rho_0$. 我们计算 r, 从而由 (65) 计算 z. 设

$$\zeta_0 = \frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}. \tag{70}$$

则一个显著性水平为 5% 的拒绝区域是

$$\sqrt{N-3}\,|z - \zeta_0| > 1.96. \tag{71}$$

一个更好的区域是

$$\sqrt{N-3}\left|z - \zeta_0 - \frac{\frac{1}{2}\rho_0}{N-1}\right| > 1.96. \tag{72}$$

b. 假设我们有来自某个总体的容量为 N_1 的样本和来自第二个总体的容量为 N_2 的样本. 我们怎样检验这两个相关系数相等的假设, $\rho_1 = \rho_2$? 由定理 4.2.5, 我们知道, 如果原假设为真, 则 $z_1 - z_2$ [其中, 对于两个样本相关系数, z_1 和 z_2 由 (65) 所定义] 是渐近正态分布的, 其均值为 0, 方差为 $1/(N_1-3) + 1/(N_2-3)$. 作为一个大小为 5% 的临界区域, 我们用

$$\frac{|z_1 - z_2|}{\sqrt{1/(N_1-3) + 1/(N_2-3)}} > 1.96. \tag{73}$$

c. 在 b 中的条件下, 假定 $\rho_1=\rho_2=\rho$. 我们怎样利用这两组样本的结果给出 ρ 的一个联合估计? 因为 z_1 和 z_2 分别有方差 $1/(N_1-3)$ 和 $1/(N_2-3)$, 我们可以用

$$\frac{(N_1-3)z_1+(N_2-3)z_2}{N_1+N_2-6} \tag{74}$$

估计 ζ 并通过 (65) 的逆将其变换为 ρ 的估计.

d. 设 r 是 N 个观测的样本相关系数. 我们怎样得到 ρ 的置信区间? 我们知道, 近似地有

$$\Pr\{-1.96\leqslant\sqrt{N-3}(z-\zeta)\leqslant 1.96\}=0.95. \tag{75}$$

由此, 我们推演出 $[-1.96/\sqrt{N-3}+z, 1.96/\sqrt{N-3}+z]$ 是 ζ 的置信区间. 从而, 我们用单调变换 $\rho=\tanh\zeta=(\mathrm{e}^{\zeta}-\mathrm{e}^{-\zeta})/(\mathrm{e}^{\zeta}+\mathrm{e}^{-\zeta})$ 得到 ρ 的区间. 因此, 95% 的置信区间是

$$\tanh(z-1.96/\sqrt{N-3})\leqslant\rho\leqslant\tanh(z+1.96/\sqrt{N-3}). \tag{76}$$

自助法用来评估样本量的变异性. 见 Efron (1982). 我们用样本相关系数来说明这个方法, 它也可以应用于本书研究的其他量.

假设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自某二元总体的样本, 这个总体不必是正态的. 自助法视这 N 个向量是大小为 N 的有限总体. 一个随机向量 $\boldsymbol{X}$ 有 (离散的) 概率

$$\Pr\{\boldsymbol{X}=\boldsymbol{x}_\alpha\}=\frac{1}{N},\quad \alpha=1,\cdots,N. \tag{77}$$

来自这个有限总体的容量为 N 的随机样本有一个概率分布, 且由这个样本计算的相关系数有一个 (离散) 的概率分布, 记为 $p_N(r)$. 自助法就是用这个分布来代替不可得到的来自母体的随机样本的相关系数的分布. 但是, 这个计算量是非常巨大的. 相反, 用来自 (77) 的大量随机样本计算得到的 r 的经验分布来估计 $p_N(r)$. Diaconis and Efron (1983) 给出了一个 $N=15$ 的例子, 他们求出了非常类似 r 真实分布的经验分布 (在这个特殊情形本质上是能得到的). 这个方法的一个优点是不需要假设母体的知识, 一个缺点是计算量很大.

4.3 偏相关系数, 条件分布

4.3.1 偏相关系数的估计

正态分布中的偏相关系数就是其条件分布里的相关系数. 2.5 节证明了: 如果 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 其中

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{pmatrix},\quad \boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \tag{1}$$

则给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件分布是 $N[\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}),\boldsymbol{\Sigma}_{11\cdot 2}]$, 其中

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}, \tag{2}$$

$$\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \tag{3}$$

给定 $\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的偏相关系数是从 $\boldsymbol{\Sigma}_{11\cdot 2}$ 用普通方法计算出的相关系数. 在本节中, 我们感兴趣的是关于这些相关系数的统计问题.

首先我们考虑基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本的估计问题. $\boldsymbol{X}^{(1)}$ (q 个分量) 的偏相关系数 $\rho_{ij\cdot q+1,\cdots,p}$ 的极大似然估计是什么? 我们知道 $\boldsymbol{\Sigma}$ 的极大似然估计是 $(1/N)\boldsymbol{A}$, 其中

$$\begin{aligned}\boldsymbol{A} &= \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' \\ &= \sum_{\alpha=1}^{N}\begin{pmatrix} \boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)} \\ \boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)} \end{pmatrix}(\boldsymbol{x}_\alpha^{(1)\prime} - \bar{\boldsymbol{x}}^{(1)\prime}, \boldsymbol{x}_\alpha^{(2)\prime} - \bar{\boldsymbol{x}}^{(2)\prime}) \\ &= \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}\end{aligned} \tag{4}$$

且 $\bar{\boldsymbol{x}} = (1/N)\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha = (\bar{\boldsymbol{x}}^{(1)\prime}, \bar{\boldsymbol{x}}^{(2)\prime})'$. 由 (2) 和 (3) 知, $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Sigma}_{11\cdot 2}$, $\boldsymbol{\beta}$ 和 $\boldsymbol{\Sigma}_{22}$ 的对应是一一的, 并且

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{\beta}\boldsymbol{\Sigma}_{22}, \tag{5}$$

$$\boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{11\cdot 2} + \boldsymbol{\beta}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}'. \tag{6}$$

应用推论 3.2.1 可以得到, 参数函数的极大似然估计是这些参数的极大似然估计的函数.

定理 4.3.1 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本, 其中如 (1) 划分 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$. 用 (4) 定义 $\boldsymbol{A}$ 且 $(\bar{\boldsymbol{x}}^{(1)\prime}, \bar{\boldsymbol{x}}^{(2)\prime}) = (1/N)\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^{(1)\prime}, \boldsymbol{x}_\alpha^{(2)\prime})$. 则 $\boldsymbol{\mu}^{(1)}$, $\boldsymbol{\mu}^{(2)}$, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{11\cdot 2}$ 和 $\boldsymbol{\Sigma}_{22}$ 的极大似然估计分别是 $\hat{\boldsymbol{\mu}}^{(1)} = \bar{\boldsymbol{x}}^{(1)}$, $\hat{\boldsymbol{\mu}}^{(2)} = \bar{\boldsymbol{x}}^{(2)}$,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}, \qquad \hat{\boldsymbol{\Sigma}}_{11\cdot 2} = \frac{1}{N}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}), \tag{7}$$

$\hat{\boldsymbol{\Sigma}}_{22} = (1/N)\boldsymbol{A}_{22}$.

依次利用推论 3.2.1, 得到 $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{22}, \sigma_{ii\cdot q+1,\cdots,p}(i=1,\cdots,q)$ 和 $\rho_{ij\cdot q+1,\cdots,p}$ $(i,j=1,\cdots,q)$ 的极大似然估计. 从而偏相关系数的极大似然估计是

$$\hat{\rho}_{ij\cdot q+1,\cdots,p} = \frac{\hat{\sigma}_{ij\cdot q+1,\cdots,p}}{\sqrt{\hat{\sigma}_{ii\cdot q+1,\cdots,p}\hat{\sigma}_{jj\cdot q+1,\cdots,p}}}, \quad i,j=1,\cdots,q, \tag{8}$$

其中 $\sigma_{ij\cdot q+1,\cdots,p}$ 是 $\hat{\boldsymbol{\Sigma}}_{11\cdot 2}$ 的第 i,j 个元素.

定理 4.3.2 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本. 给定后 $p-q$ 个分量的条件下, 前 q 个分量的偏相关系数 $\rho_{ij\cdot q+1,\cdots,p}$ 的极大似然估计是

$$\hat{\rho}_{ij\cdot q+1,\cdots,p} = \frac{a_{ij\cdot q+1,\cdots,p}}{\sqrt{a_{ii\cdot q+1,\cdots,p}a_{jj\cdot q+1,\cdots,p}}}, \quad i,j=1,\cdots,q, \tag{9}$$

其中

$$(a_{ij\cdot q+1,\cdots,p}) = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} = \boldsymbol{A}_{11\cdot 2}. \tag{10}$$

用 $r_{ij\cdot q+1,\cdots,p}$ 表示的估计 $\hat{\rho}_{ij\cdot q+1,\cdots,p}$, 称为给定 $X_{q+1},\cdots,X_p$ 时 X_i 和 X_j 的**样本偏相关系数**. 也称为 X_i 和 X_j 关于 $X_{q+1},\cdots,X_p$ 的样本偏相关系数. 注意这些可以根据 (r_{ij}) 计算.

矩阵 $\boldsymbol{A}_{11\cdot 2}$ 也可以表示为

$$\begin{aligned}\boldsymbol{A}_{11\cdot 2} &= \sum_{\alpha=1}^{N}\left[\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)} - \hat{\boldsymbol{\beta}}(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})\right]\left[\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)} - \hat{\boldsymbol{\beta}}(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})\right]' \\ &= \boldsymbol{A}_{11} - \hat{\boldsymbol{\beta}}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}'.\end{aligned} \tag{11}$$

向量 $\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)} - \hat{\boldsymbol{\beta}}(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})$ 是 $\boldsymbol{x}_\alpha^{(1)}$ 与其在 $\boldsymbol{x}_\alpha^{(2)}$ 和 1 上回归的残差. 偏相关系数是这些残差的简单相关系数. 当分布不是正态的时, 这个定义一样适用.

上述理论有两个几何解释. $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 表示 p 维空间的 N 个点. 样本回归函数

$$\boldsymbol{x}^{(1)} = \bar{\boldsymbol{x}}^{(1)} + \hat{\boldsymbol{\beta}}(\boldsymbol{x}^{(2)} - \bar{\boldsymbol{x}}^{(2)}) \tag{12}$$

是一个 $(p-q)$ 维超平面, 它是 q 个 $(p-1)$ 维超平面

$$x_i = \bar{x}_i + \sum_{j=q+1}^{p}\hat{\beta}_{ij}(x_j - \bar{x}_j), \quad i = 1,\cdots,q \tag{13}$$

的交集, 其中 x_i, x_j 是执行变量. 这里, $\hat{\beta}_{ij}$ 是 $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1} = \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$ 的一个元素. $\hat{\boldsymbol{\beta}}$ 的第 i 行是 $(\hat{\beta}_{i,q+1},\cdots,\hat{\beta}_{ip})$. 每一个 (13) 的右端是 x_i 在 $x_{q+1},\cdots,x_p$ 的最小二乘回归函数, 即如果我们将点 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 投影在 $x_i, x_{q+1},\cdots,x_p$ 的坐标超平面上, 则 (13) 是回归超平面. 坐标为

$$\begin{aligned} x_i &= \bar{x}_i + \sum_{j=q+1}^{p}\hat{\beta}_{ij}(x_j - \bar{x}_j), \quad i = 1,\cdots,q, \\ x_j &= x_{j\alpha}, \quad j = q+1,\cdots,p \end{aligned} \tag{14}$$

的点在超平面 (13) 上. $\boldsymbol{x}_\alpha$ 的第 i 个坐标和点 (14) 之差当 $i = 1,\cdots,q$ 时为 $y_{i\alpha} = x_{i\alpha} - [\bar{x}_i + \sum_{j=q+1}^{p}\hat{\beta}_{ij}(x_{j\alpha} - \bar{x}_j)]$, 在其他坐标上差为 0. 设 $\boldsymbol{y}_\alpha' = (y_{1\alpha},\cdots,y_{q\alpha})$. 这些点可以表示为 q 维空间上的 N 个点. 则 $\boldsymbol{A}_{11\cdot 2} = \sum_{\alpha=1}^{N}\boldsymbol{y}_\alpha\boldsymbol{y}_\alpha'$.

我们也可以把样本看作 N 维空间上的 p 个点 (图 4.4). 设 $\boldsymbol{u}_j = (x_{j1},\cdots,x_{jN})'$ 是第 j 个点, 且设 $\boldsymbol{\varepsilon} = (1,\cdots,1)'$ 是另外一点. 坐标为 $\bar{x}_i,\cdots,\bar{x}_i$ 的点是 $\bar{x}_i\boldsymbol{\varepsilon}$. $\boldsymbol{u}_i$ 在 $\boldsymbol{u}_{q+1},\cdots,\boldsymbol{u}_p,\boldsymbol{\varepsilon}$ 生成的超平面上的投影是

$$\hat{\boldsymbol{u}}_i = \bar{x}_i\boldsymbol{\varepsilon} + \sum_{j=q+1}^{p}\hat{\beta}_{ij}(\boldsymbol{u}_j - \bar{x}_i\boldsymbol{\varepsilon}), \tag{15}$$

这是超平面上距离 $\boldsymbol{u}_i$ 最近的点. 设 $\boldsymbol{u}_i^*$ 是向量 $\hat{\boldsymbol{u}}_i$ 和 $\boldsymbol{u}_i$ 的差, 即 $\boldsymbol{u}_i - \hat{\boldsymbol{u}}_i$, 或者等价地, 平移这个向量使得其一个端点在原点处. 向量集 $\boldsymbol{u}_1^*,\cdots,\boldsymbol{u}_q^*$ 是 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_q$ 在

超平面上正交于 $\boldsymbol{u}_{q+1},\cdots,\boldsymbol{u}_p,\boldsymbol{\varepsilon}$ 的投影. 则 $\boldsymbol{u}_i^*$ 的平方长度 (即 $\boldsymbol{u}_i$ 与 $\hat{\boldsymbol{u}}_i$ 的距离的平方) 为 $\boldsymbol{u}_i^{*\prime}\boldsymbol{u}_i^* = a_{ii\cdot q+1,\cdots,p}$. 从而 $\boldsymbol{u}_i^*$ 和 $\boldsymbol{u}_j^*$ 夹角的余弦是 $\boldsymbol{u}_i^{*\prime}\boldsymbol{u}_j^*/\sqrt{\boldsymbol{u}_i^{*\prime}\boldsymbol{u}_i^*\boldsymbol{u}_j^{*\prime}\boldsymbol{u}_j^*} = r_{ij\cdot q+1,\cdots,p}$.

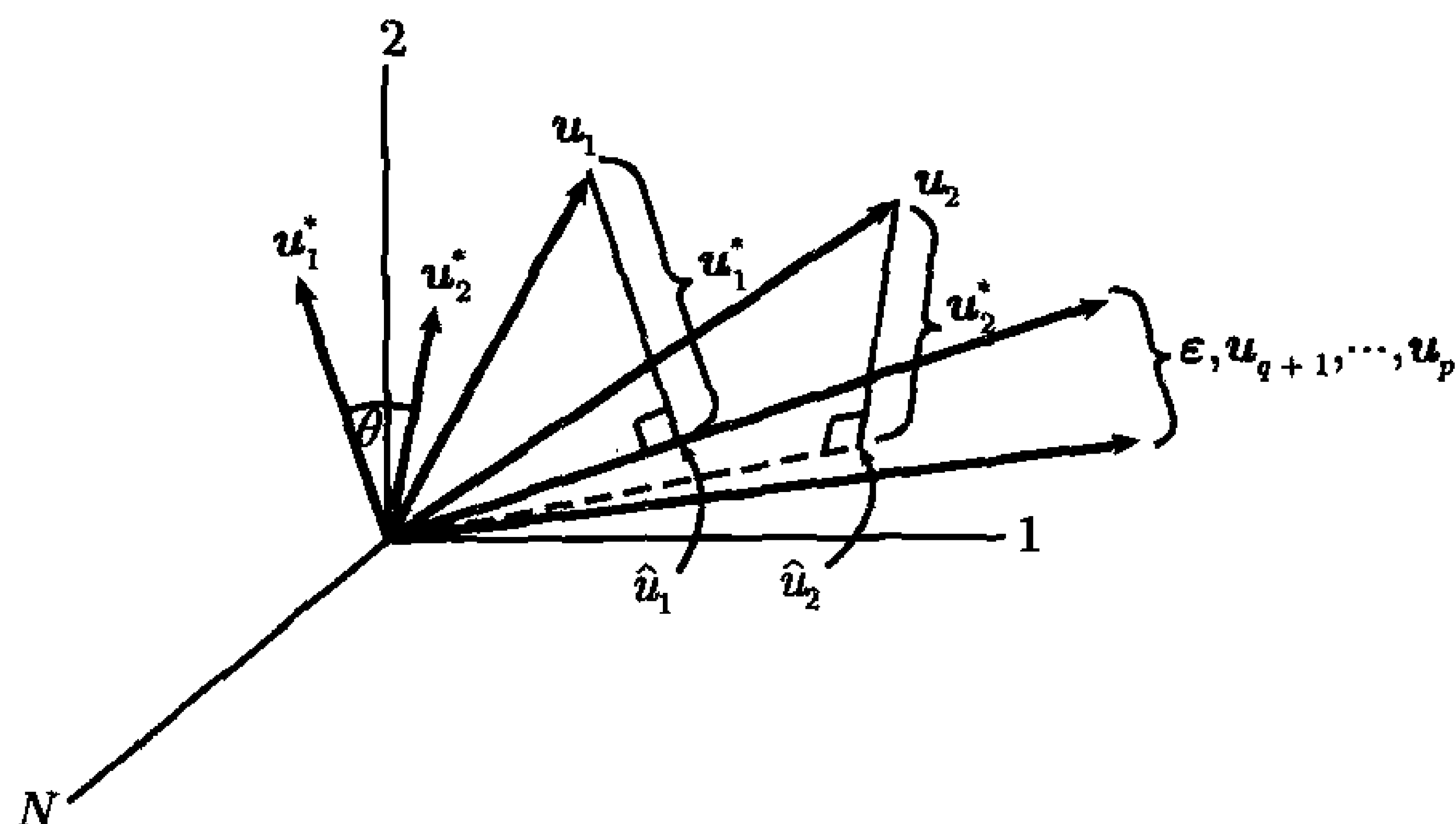


图 4.4

作为一个使用偏相关系数的例子, 我们考虑一些数据 [Hooker (1907)]: 每英亩干草产量 (X_1), 单位英担; 春季降水量 (X_2), 单位英寸; 英国地区 20 年内春季超过 42°F 的累积温度 (X_3). $\mu_i, \sigma_i(=\sqrt{\sigma_{ii}})$ 和 ρ_{ij} 的估计是

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}} = \begin{pmatrix} 28.02 \\ 4.91 \\ 594 \end{pmatrix},$$

$$\begin{pmatrix} \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ \hat{\sigma}_3 \end{pmatrix} = \begin{pmatrix} 4.42 \\ 1.10 \\ 85 \end{pmatrix}, \tag{16}$$

$$\begin{pmatrix} 1 & \hat{\rho}_{12} & \hat{\rho}_{13} \\ \hat{\rho}_{21} & 1 & \hat{\rho}_{23} \\ \hat{\rho}_{31} & \hat{\rho}_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.80 & -0.40 \\ 0.80 & 1.00 & -0.56 \\ -0.40 & -0.56 & 1.00 \end{pmatrix}.$$

我们从这些相关系数看出产量和降水量是正相关的, 产量和温度是负相关的, 降水量和温度是负相关的. 产量和温度之间显然的负相关说明了什么? 是不是高温会导致低产量, 或高温导致低降水量, 从而导致低产量? 为了解答这个问题, 我们考虑固定降水量时, 产量和温度的相关系数, 即我们用上述数据估计给定 X_2 时 X_1 和 X_3 的偏相关系数. 它是[①]

$$r_{13\cdot 2} = \frac{\hat{\sigma}_{13\cdot 2}}{\sqrt{\hat{\sigma}_{11\cdot 2}\hat{\sigma}_{33\cdot 2}}} = 0.097. \tag{17}$$

因此, 如果去掉降水量的效应, 产量和温度是正相关的. 结论是高降水量和高温度都增加干草产量, 但是大多数年份, 高降水量常伴着低温度发生, 反之亦然.

① 我们用 $\hat{\boldsymbol{\Sigma}}$ 代替 $\boldsymbol{\Sigma}$ 计算.

4.3.2　样本偏相关系数的分布

为了检验一个关于总体偏相关系数的假设, 我们需要样本偏相关系数的分布. 用从 A 计算相关系数的相同方法从 $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ (如定理 4.3.1 指出的) 计算偏相关系数. 为了获得样本偏相关系数的分布, 我们证明了 A 和 $\sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha$ 同分布, 其中 $Z_1, \cdots, Z_{N-1}$ 独立服从分布 $N(0, \Sigma)$ 且和 $\bar{X}$ 独立 (定理 3.3.2). 这里, 我们要证明 $A_{11\cdot 2}$ 和 $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U'_\alpha$ 同分布, 其中 $U_1, \cdots, U_{N-1-(p-q)}$ 独立服从分布 $N(0, \Sigma_{11\cdot 2})$ 且和 $\hat{\boldsymbol{\beta}}$ 独立. 样本偏相关系数的分布将从 $A_{11\cdot 2}$ 的分布特征中得到. 我们用一般形式表述这个定理, 在第 8 章详细讨论回归时将会利用到这个定理. 紧接的推论把这个定理应用到 $A_{11\cdot 2}$, 并用残差来表示.

定理 4.3.3　假设 $Y_1, \cdots, Y_m$ 独立于分布为 $N(\Gamma w_\alpha, \Phi)$ 的 Y_α, 其中 w_α 是一个 r 维向量. 设 $H = \sum_{\alpha=1}^m w_\alpha w'_\alpha$, 假定 H 非奇异, $G = \sum_{\alpha=1}^m Y_\alpha w'_\alpha H^{-1}$,

$$C = \sum_{\alpha=1}^m (Y_\alpha - Gw_\alpha)(Y_\alpha - Gw_\alpha)' = \sum_{\alpha=1}^m Y_\alpha Y'_\alpha - GHG'. \tag{18}$$

则 C 和 $\sum_{\alpha=1}^{m-r} U_\alpha U'_\alpha$ 同分布, 其中 $U_1, \cdots, U_{m-r}$ 独立服从分布 $N(0, \Phi)$ 且独立于 G.

证明　$Y = (Y_1, \cdots, Y_m)$ 的行是 m 维空间的随机向量, $W = (w_1, \cdots, w_m)$ 的行是这个空间的固定向量. 这个证明的思想是旋转坐标轴使得最后 r 个轴在 W 的行向量生成的空间里. 设 $E_2 = FW$, 其中 F 是一个满足 $FHF' = I$ 的方阵. 则

$$\begin{aligned} E_2E'_2 &= FWW'F' = F\sum_{\alpha=1}^m w_\alpha w'_\alpha F' \\ &= FHF' = I. \end{aligned} \tag{19}$$

因此, E_2 的 m 维行向量相互正交且有单位长度. 可以找到一个 $(m-r)\times m$ 的矩阵 E_1, 使得

$$E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \tag{20}$$

是正交的. (见附录, 引理 A.4.2.) 现在设 $U = YE'$ (即 $U_\alpha = \sum_{\beta=1}^m e_{\alpha\beta} Y_\beta$). 由定理 3.3.1, $U = (U_1, \cdots, U_m)$ 的列是独立的和正态分布的, 每一个列向量的协方差阵是 Φ. 由 E 的正交性, 可知均值是

$$\begin{aligned} \mathrm{E}(U) &= \mathrm{E}(YE') = \Gamma WE' \\ &= \Gamma F^{-1}E_2(E'_1 E'_2) \\ &= (0 \quad \Gamma F^{-1}) \end{aligned} \tag{21}$$

为了完成定理的证明, 我们需要说明 C 变换成 $\sum_{\alpha=1}^{m-r} U_\alpha U'_\alpha$. 我们有

$$\sum_{\alpha=1}^m Y_\alpha Y'_\alpha = YY' = UEE'U' = UU' = \sum_{\alpha=1}^m U_\alpha U'_\alpha. \tag{22}$$

注意到

$$G = YW'H^{-1} = UEE_2'(F^{-1})'F'F \tag{23}$$
$$= U\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} E_2'F$$
$$= U\begin{pmatrix} 0 \\ I \end{pmatrix} F = U^{(2)}F,$$

其中 $U^{(2)} = (U_{m-r+1}, \cdots, U_m)$. 则

$$GHG' = U^{(2)}FHF'U^{(2)'} = U^{(2)}U^{(2)'} = \sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha'. \tag{24}$$

因此 C 是

$$\sum_{\alpha=1}^{m} Y_\alpha Y_\alpha' - GHG' = \sum_{\alpha=1}^{m} U_\alpha U_\alpha' - \sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha' = \sum_{\alpha=1}^{m-r} U_\alpha U_\alpha'. \tag{25}$$

定理得证. ∎

当 $\boldsymbol{\Gamma} = \mathbf{0}$ 时, 由上面所述, $\mathrm{E}(U) = \mathbf{0}$, 并且我们得到下面推论.

推论 4.3.1 如果 $\boldsymbol{\Gamma} = \mathbf{0}$, 则定理 4.3.3 中定义的矩阵 GHG' 和 $\sum_{\alpha=m-r+1}^{m} U_\alpha U_\alpha'$ 同分布, 其中 $U_{m-r+1}, \cdots, U_m$ 独立服从分布 $N(\mathbf{0}, \boldsymbol{\Phi})$.

我们现在用相同形式求 $A_{11\cdot 2}$ 的分布. 定理 3.3.1 证明了 A 与 $\sum_{\alpha=1}^{N-1} Z_\alpha Z_\alpha'$ 同分布, 其中 $Z_1, \cdots, Z_{N-1}$ 独立服从分布 $N(\mathbf{0}, \boldsymbol{\Sigma})$. 设 Z_α 划分成分别为 q 维和 $p-q$ 维的子向量:

$$Z_\alpha = \begin{pmatrix} Z_\alpha^{(1)} \\ Z_\alpha^{(2)} \end{pmatrix}. \tag{26}$$

则 $A_{ij} = \sum_{\alpha=1}^{N} Z_\alpha^{(i)} Z_\alpha^{(j)'}$. 由引理 4.2.1, 在 $Z_1^{(2)} = z_1^{(2)}, \cdots, Z_{N-1}^{(2)} = z_{N-1}^{(2)}$ 条件下, 随机向量 $Z_1^{(1)}, \cdots, Z_{N-1}^{(1)}$ 是独立分布的, $Z_\alpha^{(1)}$ 服从分布 $N(\boldsymbol{\beta} z_\alpha^{(2)}, \boldsymbol{\Sigma}_{11\cdot 2})$, 其中 $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ 和 $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. 现在我们应用定理 4.3.3, 取 $Z_\alpha^{(1)} = Y_\alpha, z_\alpha^{(2)} = w_\alpha, N-1 = m, p-q = r, \boldsymbol{\beta} = \boldsymbol{\Gamma}, \boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Phi}, A_{11} = \sum_{\alpha=1}^{N-1} Y_\alpha Y_\alpha', A_{12}A_{22}^{-1} = G, A_{22} = H$. 我们发现给定 $Z_\alpha^{(2)} = z_\alpha^{(2)}, \alpha = 1, \cdots, N-1$ 时, $(A_{12}A_{22}^{-1})A_{22}(A_{22}^{-1}A_{12}') = A_{11\cdot 2}$ 的条件分布就是 $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U_\alpha'$ 的分布, 其中 $U_1, \cdots, U_{N-1-(p-q)}$ 独立服从分布 $N(\mathbf{0}, \boldsymbol{\Sigma}_{11\cdot 2})$. 因为此分布不依赖 $\{z_\alpha^{(2)}\}$, 我们得到下面的定理.

定理 4.3.4 矩阵 $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$ 和 $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U_\alpha'$ 同分布, 其中 $U_1, \cdots, U_{N-1-(p-q)}$ 独立服从分布 $N(\mathbf{0}, \boldsymbol{\Sigma}_{11\cdot 2})$, 且独立于 A_{12} 和 A_{22}.

推论 4.3.2 如果 $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ ($\boldsymbol{\beta} = \mathbf{0}$), 则 $A_{11\cdot 2}$ 和 $\sum_{\alpha=1}^{N-1-(p-q)} U_\alpha U_\alpha'$ 同分布, 且 $A_{12}A_{22}^{-1}A_{21}$ 和 $\sum_{\alpha=N-(p-q)}^{N-1} U_\alpha U_\alpha'$ 同分布, 其中 $U_1, \cdots, U_{N-1}$ 独立服从分布 $N(\mathbf{0}, \boldsymbol{\Sigma}_{11\cdot 2})$.

于是可知, 基于 N 个观测的 $r_{ij\cdot q+1,\cdots,p}$ 的分布等于, 基于来自母体相关系数值为 $\rho_{ij\cdot q+1,\cdots,p}$ 的 $N-(p-q)$ 个观测的简单相关系数.

定理 4.3.5 如果基于来自相关系数为 ρ_{ij} 的正态分布的容量为 N 的样本的 r_{ij} 的 cdf 被表示为 $F(r|N,\rho_{ij})$, 则基于来自偏相关系数为 $\rho_{ij\cdot q+1,\cdots,p}$ 的正态分布的容量为 N 的样本的样本偏相关系数 $r_{ij\cdot q+1,\cdots,p}$ 的 cdf 为 $F[r|N-(p-q),\rho_{ij}\cdot q+1,\cdots,p]$.

这个分布是 Fisher (1924) 得到的.

4.3.3 偏相关系数的假设检验和置信区域

因为基于来自总体相关系数 $\rho_{ij\cdot q+1,\cdots,p}$ 为某特定值 (如 ρ) 的分布的容量为 N 的样本, 其样本偏相关系数 $r_{ij\cdot q+1,\cdots,p}$ 的分布与基于来自总体相关系数为 ρ 的容量为 $N-(p-q)$ 的样本的简单相关系数 r 的分布相同, 故而所有对简单相关系数的统计推断方法都可以用在偏相关系数上. 对偏相关系数的方法是一样的, 除了把 N 换成 $N-(p-q)$ 外. 为了说明这个规则, 我们给出两个例子.

例 1 假设在容量为 N 的样本基础上, 我们想求 $\rho_{ij\cdot q+1,\cdots,p}$ 的置信区间. 样本偏相关系数是 $r_{ij\cdot q+1,\cdots,p}$. 方法是利用 $N-(p-q)$ 的 David 图表. 在 4.3.1 节最后的例子中, 我们想求 $\rho_{12\cdot 3}$ 的置信度为 0.95 的置信区间. 样本偏相关系数是 $r_{12\cdot 3}=0.759$. 我们利用 $N-(p-q)=20-1=19$ 的图表 (或表). 这个区间是 $0.50<r_{12\cdot 3}<0.88$.

例 2 假设在容量为 N 的样本基础上, 我们对 $\rho_{ij\cdot q+1,\cdots,p}=\rho_0$ 的一个近似显著性检验及其双边备择应用 Fisher z. 我们设

$$z=\frac{1}{2}\ln\frac{1+r_{ij\cdot q+1,\cdots,p}}{1-r_{ij\cdot q+1,\cdots,p}},\quad \zeta_0=\frac{1}{2}\ln\frac{1+\rho_0}{1-\rho_0}. \tag{27}$$

则比较 $\sqrt{N-(p-q)-3}(z-\zeta_0)$ 和标准正态分布的分位数. 在 4.3.1 节最后的例子中, 我们将以显著性水平 0.05 检验假设 $\rho_{13\cdot 2}=0$. 则 $\zeta_0=0$ 和 $\sqrt{20-1-3}(0.0973)=0.3892$. 这个值显然是不显著的 ($|0.3892|<1.96$), 因而由这些数据不能拒绝原假设.

为了回答当两个变量 x_1 和 x_2 都与一个向量 $\boldsymbol{x}^{(2)}=(x_3,\cdots,x_p)$ 相关时, 他们是否相关的问题, 有两个方法可以利用. 一是考虑 x_1 在 x_2 和 $\boldsymbol{x}^{(2)}$ 上的回归并检验 x_1 在 x_2 上的回归是否为 0. 另一个方法是检验 $\rho_{12\cdot 3,\cdots,p}$ 是否为 0. 习题 4.43 ~ 4.47 证明这些方法可以得到相同的检验.

4.4 多重相关系数

4.4.1 多重相关系数的估计

2.5 节定义了一个变量和一组变量的总体多重相关系数. 为了方便, 我们将考虑 X_1 和向量 $\boldsymbol{X}^{(2)}=(X_2,\cdots,X_p)'$ 的多重相关系数, 我们不需要在 R 上用下标.

这些变量通常是可数个的, 因此所求多重相关系数就是这一个 (忽略任何不相关变量). 总体的多重相关系数是

$$\bar{R}=\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}{\sqrt{\sigma_{11}\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}}=\sqrt{\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}{\sigma_{11}}}=\sqrt{\frac{\boldsymbol{\sigma}'_{(1)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}}{\sigma_{11}}}, \tag{1}$$

其中 $\boldsymbol{\beta}$, $\boldsymbol{\sigma}_{(1)}$ 和 $\boldsymbol{\Sigma}_{22}$ 定义为

$$\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11} & \boldsymbol{\sigma}'_{(1)}\\ \boldsymbol{\sigma}_{(1)} & \boldsymbol{\Sigma}_{22}\end{pmatrix}, \tag{2}$$

$$\boldsymbol{\beta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}. \tag{3}$$

给定样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ $(N>p)$, 我们用 $\boldsymbol{S}=[N/(N-1)]\hat{\boldsymbol{\Sigma}}$ 或者

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{N}\boldsymbol{A}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'=\begin{pmatrix}\hat{\sigma}_{11} & \hat{\boldsymbol{\sigma}}'_{(1)}\\ \hat{\boldsymbol{\sigma}}_{(1)} & \hat{\boldsymbol{\Sigma}}_{22}\end{pmatrix} \tag{4}$$

估计 $\boldsymbol{\Sigma}$, 用 $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\sigma}}_{(1)}=\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}$ 估计 $\boldsymbol{\beta}$. 我们定义样本多重相关系数为

$$R=\sqrt{\frac{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Sigma}}_{22}\hat{\boldsymbol{\beta}}}{\hat{\sigma}_{11}}}=\sqrt{\frac{\hat{\boldsymbol{\sigma}}'_{(1)}\hat{\boldsymbol{\Sigma}}_{22}^{-1}\hat{\boldsymbol{\sigma}}_{(1)}}{\hat{\sigma}_{11}}}=\sqrt{\frac{\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11}}}. \tag{5}$$

因为我们可以定义 $\bar{R}$, $\boldsymbol{\sigma}_{(1)}$, $\boldsymbol{\Sigma}_{22}$ 为 $\boldsymbol{\Sigma}$ 的一一变换, 则由推论 3.2.1 知这是 $\bar{R}$ 的极大似然估计. R 的另一个表达式 [见 2.5 节 (16)] 来自

$$1-R^2=\frac{|\hat{\boldsymbol{\Sigma}}|}{\hat{\boldsymbol{\sigma}}_{11}|\hat{\boldsymbol{\Sigma}}_{22}|}=\frac{|\boldsymbol{A}|}{a_{11}|\boldsymbol{A}_{22}|}. \tag{6}$$

样本的量 R 和 $\hat{\boldsymbol{\beta}}$ 与总体的 $\bar{R}$ 和 $\boldsymbol{\beta}$ 有相似的性质. 我们对定理 2.5.2、定理 2.5.3 和定理 2.5.4 进行类推. 设 $\hat{x}_{1\alpha}=\bar{x}_1+\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$, 并且 $x_{1\alpha}^*=x_{1\alpha}-\hat{x}_{1\alpha}$ 为残差.

定理 4.4.1 残差 $x_{1\alpha}^*$ 与 $\boldsymbol{x}_\alpha^{(2)}$ 的分量在样本中不相关, $\alpha=1,\cdots,N$. 对每一个向量 $\boldsymbol{a}$,

$$\sum_{\alpha=1}^{N}[x_{1\alpha}-\bar{x}_1-\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2\leqslant\sum_{\alpha=1}^{N}[x_{1\alpha}-\bar{x}_1-\boldsymbol{a}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2. \tag{7}$$

$x_{1\alpha}$ 和 $\boldsymbol{a}'\boldsymbol{x}_\alpha^{(2)}(\alpha=1,\cdots,N)$ 的样本相关系数在 $\boldsymbol{a}=\hat{\boldsymbol{\beta}}$ 时达到极大值, 且这个极大相关系数是 R.

证明 因为残差的样本均值是 0, 则 $x_{1\alpha}^*$ 和 $\boldsymbol{x}_\alpha^{(2)}$ 的样本协方差向量等比于

$$\sum_{\alpha=1}^{N}[x_{1\alpha}-\bar{x}_1-\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})](\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})'=\boldsymbol{a}'_{(1)}-\hat{\boldsymbol{\beta}}'\boldsymbol{A}_{22}=\boldsymbol{0}. \tag{8}$$

(7) 的右端可以写成左端加上

$$\sum_{\alpha=1}^{N}[(\hat{\boldsymbol{\beta}}-\boldsymbol{a})'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2 \tag{9}$$

$$=(\hat{\boldsymbol{\beta}}-\boldsymbol{a})'\sum_{\alpha=1}^{N}(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})'(\hat{\boldsymbol{\beta}}-\boldsymbol{a}),$$

这一项等于 0 当且仅当 $\boldsymbol{a}=\hat{\boldsymbol{\beta}}$. 为了证明第三个结论, 我们考虑满足 $\sum_{\alpha=1}^{N}[\boldsymbol{a}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2=\sum_{\alpha=1}^{N}[\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2$ 的向量 $\boldsymbol{a}$, 当线性函数乘以一个正常数时, 相关系数是不变的. 从 (7) 我们得到

$$\begin{aligned}&a_{11}-2\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})+\sum_{\alpha=1}^{N}\left[\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})\right]^2\\&\leqslant a_{11}-2\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)\boldsymbol{a}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})+\sum_{\alpha=1}^{N}\left[\boldsymbol{a}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})\right]^2,\end{aligned}\tag{10}$$

我们可以从中得到

$$\begin{aligned}\frac{\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{a}}{\sqrt{a_{11}}\sqrt{\sum_{\alpha=1}^{N}\left[\boldsymbol{a}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})\right]^2}}&\leqslant\frac{\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})'\hat{\boldsymbol{\beta}}}{\sqrt{a_{11}}\sqrt{\sum_{\alpha=1}^{N}\left[\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})\right]^2}}\\&=\frac{\boldsymbol{a}_{(1)}'\hat{\boldsymbol{\beta}}}{\sqrt{a_{11}}\sqrt{\hat{\boldsymbol{\beta}}'\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}}},\end{aligned}\tag{11}$$

即 (5). ■

因此, $\bar{x}_1+\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 是样本中 $x_{1\alpha}$ 的最优线性预测, 且 $\hat{\boldsymbol{\beta}}'\boldsymbol{x}_{\alpha}^{(2)}$ 是和 $x_{1\alpha}$ 有极大样本相关系数的 $\boldsymbol{x}_{\alpha}^{(2)}$ 的线性函数. 离差平方和 [(7) 的左端] 极小值是

$$\begin{aligned}\sum_{\alpha=1}^{N}[x_{1\alpha}-\bar{x}_1-\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})]^2&=a_{11}-\hat{\boldsymbol{\beta}}'\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}\\&=a_{11}-\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}\\&=a_{11\cdot2},\end{aligned}\tag{12}$$

如 4.3 节定义, 取 $q=1$. $\sigma_{11\cdot2}$ 的极大似然估计是 $\hat{\sigma}_{11\cdot2}=a_{11\cdot2}/N$. 则有

$$\hat{\sigma}_{11\cdot2}=(1-R^2)\hat{\sigma_{11}}.\tag{13}$$

因此, $1-R^2$ 度量了用残差减少方差的比例. 我们说 R^2 是由 $\boldsymbol{x}^{(2)}$ 解释的部分方差. R^2 越大, 用 $\boldsymbol{x}^{(2)}$ 里解释变量减少的方差越多.

在 p 维空间里, $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 表示 N 个点. 样本回归函数 $x_1=\bar{x}_1+\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 是 $(p-1)$ 维的超平面, 它使得那些点到其上的平方离差极小, 这些离差可以在 x_1 方向上计算. 这个超平面通过点 $\bar{\boldsymbol{x}}$.

在 N 维空间, $(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 的行表示 p 个点. 第 α 个分量为 $x_{i\alpha}-\bar{x}_i$ 的 N 维向量是第 α 个分量为 $x_{i\alpha}$ 的向量在与等角线正交的平面上的投影. 我们有 p 个这样的向量, $\boldsymbol{a}'(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 是由最后 $(p-1)$ 个向量张成的超平面里的一个向量的第

α 个分量. 因为 (7) 的右端是第一个向量和最后 $p-1$ 个向量的线性组合的平方距离, $\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 是极小化这个距离的向量的一个分量. (8) 的解释是第 α 个分量是 $(x_{1\alpha}-\bar{x}_1)-\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 的向量与最后 $p-1$ 个向量分别正交. 因此, 第 α 个分量为 $\hat{\boldsymbol{\beta}}'(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 的向量是第一个向量在超平面上的投影. 见图 4.5. 这个投影向量的平方长度是

$$\sum_{\alpha=1}^{N}\left[\hat{\boldsymbol{\beta}}'\left(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)}\right)\right]^2=\hat{\boldsymbol{\beta}}'\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}=\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}, \tag{14}$$

且第一个向量的平方长度是 $\sum_{\alpha=1}^{N}(x_{1\alpha}-\bar{x}_1)^2=a_{11}$. 因此 R 是第一个向量与其投影的夹角的余弦.

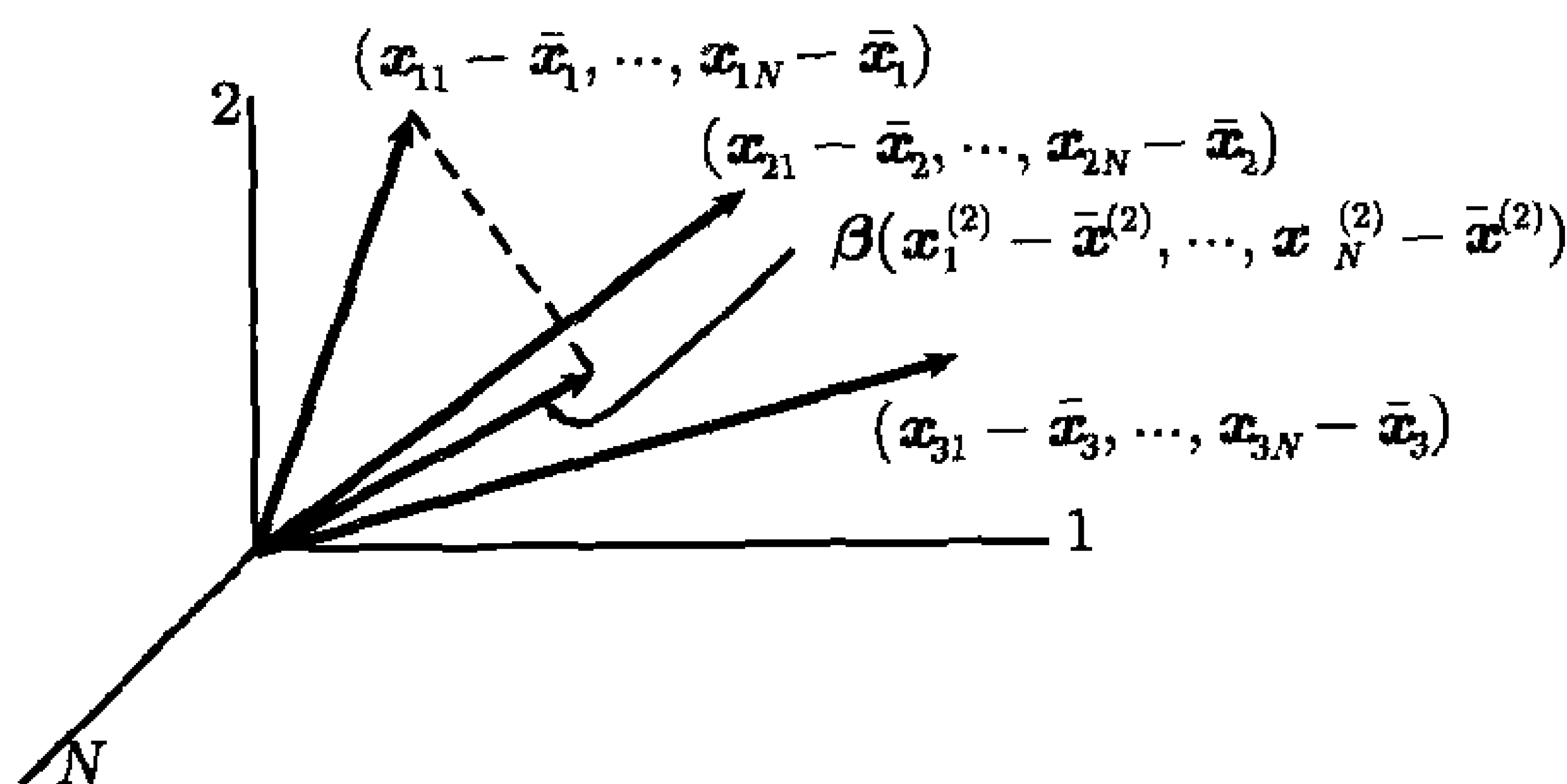


图 4.5

在 3.2 节, 我们知道简单相关系数是两个有关向量 (在正交于等角线的平面上) 的夹角余弦. R 是 $x_{1\alpha}$ 与 $\boldsymbol{x}_\alpha^{(2)}$ 的分量的线性组合之间的相关系数的极大值, 这一性质对应于一个几何性质, 即 R 是一个分量为 $x_{1\alpha}-\bar{x}_1$ 的向量与其他 $p-1$ 个向量张成的超平面上的一个向量的最小夹角的余弦.

这个几何解释是从正交于等角线的 $(N-1)$ 维超平面上的向量来讲的. 3.3 节证明了这个超平面上的向量 $(x_{i1}-\bar{x}_i,\cdots,x_{iN}-\bar{x}_i)$ 可以指定为 $(z_{i1},\cdots,z_{i,N-1})$, 其中 $z_{i\alpha}$ 是关于这个超平面里的一个 $(N-1)$ 维坐标系统的坐标. 已经证明新坐标是由原来的坐标通过变换 $z_{i\alpha}=\sum_{\beta=1}^{N}b_{\alpha\beta}x_{i\beta}(\alpha=1,\cdots,N)$ 而得, 其中 $\boldsymbol{B}=(b_{\alpha\beta})$ 是最后一行为 $(1/\sqrt{N},\cdots,1/\sqrt{N})$ 的正交矩阵. 则

$$a_{ij}=\sum_{\alpha=1}^{N}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j)=\sum_{\alpha=1}^{N-1}z_{i\alpha}z_{j\alpha}. \tag{15}$$

为了方便, 称由 $z_{i\alpha}$ 定义的多重相关系数为**不减去均值的多重相关系数**.

总体多重相关系数 $\bar{R}$ 本质上是在 X_1 的位移变换和比例变换以及 $\boldsymbol{X}^{(2)}$ 的非奇异线性变换 (即变换 $X_1^*=cX_1+d, \boldsymbol{X}^{(2)*}=\boldsymbol{C}\boldsymbol{X}^{(2)}+\boldsymbol{d}$) 下, 参数 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的不变的唯一函数. 类似地, 样本多重相关系数 R 本质上只是 $\bar{\boldsymbol{x}}$ 和 $\hat{\boldsymbol{\Sigma}}$ ($\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的充分统计量集) 的函数, 它是在这些变换下不变的. 就像样本的简单相关系数 r 是两个标量变量的关联测度, 样本的多重相关系数 R 是一个标量变量和一个向量变量的关联测度.

4.4.2 当总体多重相关系数为零时样本多重相关系数的分布

从 (5) 中, 我们得到

$$R^2 = \frac{\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11}}, \tag{16}$$

则

$$1-R^2 = 1-\frac{\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11}} = \frac{a_{11}-\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11}} = \frac{a_{11\cdot 2}}{a_{11}}, \tag{17}$$

$$\frac{R^2}{1-R^2} = \frac{\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}}{a_{11\cdot 2}}. \tag{18}$$

取 $q=1$, 推论 4.3.2 说明当 $\boldsymbol{\beta}=\mathbf{0}$ 时, 即当 $\bar{R}=0$ 时, $a_{11\cdot 2}$ 与 $\sum_{\alpha=1}^{N-p}V_\alpha^2$ 同分布且 $\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}$ 和 $\sum_{\alpha=N-p+1}^{N-1}V_\alpha^2$ 同分布, 其中 $V_1,\cdots,V_{N-1}$ 独立服从分布 $N(0,\sigma_{11\cdot 2})$. 则 $a_{11\cdot 2}/\sigma_{11\cdot 2}$ 和 $\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}$ 独立, 分别服从自由度为 $N-p$ 和 $p-1$ 的 χ^2 分布. 因此

$$\begin{aligned}\frac{R^2}{1-R^2}\cdot\frac{N-p}{p-1} &= \frac{\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot 2}}{a_{11\cdot 2}/\sigma_{11\cdot 2}}\cdot\frac{N-p}{p-1} \\ &= \frac{\chi_{p-1}^2}{\chi_{N-p}^2}\cdot\frac{N-p}{p-1} \\ &= F_{p-1,N-p}\end{aligned} \tag{19}$$

服从自由度为 $p-1$ 和 $N-p$ 的 F 分布. F 的密度是

$$\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(N-p)]}\left(\frac{p-1}{N-p}\right)^{\frac{1}{2}(p-1)} f^{\frac{1}{2}(p-1)-1}\left(1+\frac{p-1}{N-p}f\right)^{-\frac{1}{2}(N-1)}. \tag{20}$$

因此

$$R = \sqrt{\frac{\frac{p-1}{N-p}F_{p-1,N-p}}{1+\frac{p-1}{N-p}F_{p-1,N-p}}} \tag{21}$$

的密度是

$$2\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(p-1)]\Gamma[\frac{1}{2}(N-p)]}R^{p-2}(1-R^2)^{\frac{1}{2}(N-p)-1}, \qquad 0\leqslant R\leqslant 1. \tag{22}$$

定理 4.4.2 对于一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本, 设 R 是 X_1 和 $\boldsymbol{X}^{(2)\prime}=(X_2,\cdots,X_p)$ 的样本多重相关系数 [由 (5) 定义]. 如果 $\bar{R}=0$ [即如果 $(\sigma_{12},\cdots,\sigma_{1p})'=\mathbf{0}=\boldsymbol{\beta}$], 则 $[R^2/(1-R^2)]\cdot[(N-p)/(p-1)]$ 服从自由度为 $p-1$ 和 $N-p$ 的 F 分布.

需要注意 $p-1$ 是 $\boldsymbol{X}^{(2)}$ 的分量个数且 $N-p=N-(p-1)-1$. 如果多重相关系数是一个分量 X_i 和 q 个其他分量之间的, 则这些数是 q 和 $N-q-1$.

注意到 $R^2/(1-R^2)$ 是在回归 (或最小二乘) 理论中出现的检验假设 X_1 在 $X_2,\cdots,X_p$ 上回归为零的量.

如果 $\bar{R}\neq 0$, 则 R 的分布很难得到. 我们在 4.4.3 节求这个分布.

现在让我们考虑基于一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本检验假设 $H:\bar{R}=0$ 的统计问题. [$\bar{R}$ 是 X_1 和 $(X_2,\cdots,X_p)$ 的总体多重相关系数.] 因为 $\bar{R}\geqslant 0$, 则备择假设是 $\bar{R}>0$.

我们来求这个假设的似然比检验. 这个似然函数是

$$L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)=\frac{1}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{\Sigma}^*|^{\frac{1}{2}N}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}^*)\right]. \tag{23}$$

观测值已知, L 是不定元 $\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*$ 的一个函数. 设 ω 是参数空间 Ω 中的原假设所在的区域. 似然比准则是

$$\lambda=\frac{\max_{\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*\in\omega}L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)}{\max_{\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*\in\Omega}L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)}. \tag{24}$$

这里 Ω 是 $\boldsymbol{\mu}^*$ 和正定的 $\boldsymbol{\Sigma}^*$ 的空间, ω 是这个空间里满足 $\bar{R}=\sqrt{\boldsymbol{\sigma}'_{(1)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}}/\sqrt{\sigma_{11}}=0$, 即 $\boldsymbol{\sigma}'_{(1)}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}=\boldsymbol{0}$ 的区域. 因为 $\boldsymbol{\Sigma}_{22}^{-1}$ 是正定的, 所以这个条件等价于 $\boldsymbol{\sigma}_{(1)}=\boldsymbol{0}$. 在 Ω 上 $L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)$ 的极大值出现在 $\boldsymbol{\mu}^*=\hat{\boldsymbol{\mu}}=\bar{\boldsymbol{x}}$ 和 $\boldsymbol{\Sigma}^*=\hat{\boldsymbol{\Sigma}}=(1/N)\boldsymbol{A}=(1/N)\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$ 且极大值是

$$\max_{\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*\in\Omega}L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)=\frac{N^{-\frac{1}{2}pN}\mathrm{e}^{\frac{1}{2}pN}}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{A}|^{\frac{1}{2}N}}. \tag{25}$$

在 ω 上, 极大似然函数是

$$\begin{aligned}L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*|\boldsymbol{\sigma}^*_{(1)}=\boldsymbol{0})=&\frac{1}{(2\pi)^{\frac{1}{2}N}\sigma_{11}^{*\frac{1}{2}N}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(x_{1\alpha}-\mu_1^*)^2/\sigma_{11}^*\right] \\ &\cdot\frac{1}{(2\pi)^{\frac{1}{2}(p-1)N}|\boldsymbol{\Sigma}_{22}^*|^{\frac{1}{2}N}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^{(2)}-\boldsymbol{\mu}^{(2)*})'\boldsymbol{\Sigma}_{22}^{*-1}(\boldsymbol{x}_\alpha^{(2)}-\boldsymbol{\mu}^{(2)*})\right].\end{aligned} \tag{26}$$

第一个因子在 $\mu_1^*=\hat{\mu}_1=\bar{x}_1$ 和 $\sigma_{11}^*=\hat{\sigma}_{11}=(1/N)a_{11}$ 处极大化且第二个因子在 $\boldsymbol{\mu}^{(2)*}=\hat{\boldsymbol{\mu}}^{(2)}=\bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{\Sigma}_{22}^*=\hat{\boldsymbol{\Sigma}}_{22}=(1/N)\boldsymbol{A}_{22}$ 处极大化. 极大化的函数值是

$$\max_{\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*\in\omega}L(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*)=\frac{N^{\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}N}}{(2\pi)^{\frac{1}{2}N}a_{11}^{\frac{1}{2}N}}\cdot\frac{N^{\frac{1}{2}(p-1)N}\mathrm{e}^{-\frac{1}{2}(p-1)N}}{(2\pi)^{\frac{1}{2}(p-1)N}|\boldsymbol{A}_{22}|^{\frac{1}{2}N}}. \tag{27}$$

因此似然比准则是 [见 (6)]

$$\chi=\frac{|\boldsymbol{A}|^{\frac{1}{2}N}}{a_{11}^{\frac{1}{2}N}|\boldsymbol{A}_{22}|^{\frac{1}{2}N}}=(1-R^2)^{\frac{1}{2}N}. \tag{28}$$

这个似然比检验由临界区域 $\lambda<\lambda_0$ 构成, 其中 λ_0 的选择使得这个不等式当 $\bar{R}=0$ 时的概率是显著性水平 α. 一个等价的检验是

$$1-\lambda^{2/N}=R^2>1-\lambda_0^{2/N}. \tag{29}$$

因为 $[R^2/(1-R^2)][(N-p)/(p-1)]$ 是 R 的一个单调函数, 则一个等价的检验是这个比大于某一个常数. 当 $\bar{R}=0$ 时, 这个比服从 $F_{p-1,N-p}$ 分布. 因此, 临界区域是

$$\frac{R^2}{1-R^2}\cdot\frac{N-p}{p-1}>F_{p-1,N-p}(\alpha),\tag{30}$$

其中 $F_{p-1,N-p}(\alpha)$ 是相应于 α 显著性水平的 (上) 分位数.

定理 4.4.3　给定来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, (30) 给出了假设 $\bar{R}=0$ 的显著性水平为 α 的似然比检验, 其中 $\bar{R}$ 是 X_1 和 $(X_2,\cdots,X_p)$ 的总体多重相关系数, R 是 (5) 定义的样本多重相关系数.

例如, 考虑 4.3.1 节最后给出的数据. 样本多重相关系数可从下式求得:

$$1-R^2=\frac{\begin{vmatrix}1 & r_{12} & r_{13}\\ r_{21} & 1 & r_{23}\\ r_{31} & r_{32} & 1\end{vmatrix}}{\begin{vmatrix}1 & r_{23}\\ r_{32} & 1\end{vmatrix}}=\frac{\begin{vmatrix}1.00 & 0.80 & -0.40\\ 0.80 & 1.00 & -0.56\\ -0.40 & -0.56 & 1.00\end{vmatrix}}{\begin{vmatrix}1.00 & -0.56\\ -0.56 & 1.00\end{vmatrix}}=0.357.\tag{31}$$

因此, R 是 0.802. 如果我们想以水平 0.01 检验干草产量独立于春季降水量和温度的假设, 则我们比较观测到的 $[R^2/(1-R^2)][(20-3)/(3-1)]=15.3$ 和 $F_{2,17}(0.01)=6.11$, 从而求得结论是显著的, 即拒绝原假设.

X_1 和 $(X_2,\cdots,X_p)=\boldsymbol{X}^{(2)\prime}$ 的独立性检验等价于检验如果 X_1 在 $\boldsymbol{x}^{(2)}$ 上的回归 (即给定 $X_2=x_2,\cdots,X_p=x_p$ 时 X_1 的条件期望) 是 $\mu_1+\boldsymbol{\beta}'(\boldsymbol{X}^{(2)}-\boldsymbol{\mu}^{(2)})$, 则回归系数向量是 $\mathbf{0}$. 这里 $\hat{\boldsymbol{\beta}}=\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}$ 是 $\boldsymbol{\beta}$ 的普通最小二乘估计, 其均值为 $\boldsymbol{\beta}$, 协方差阵为 $\sigma_{11\cdot2}\boldsymbol{A}_{22}^{-1}$ ($\boldsymbol{X}_\alpha^{(2)}$ 是固定的), 且 $a_{11\cdot2}/(N-p)$ 是 $\sigma_{11\cdot2}$ 的普通估计. 因此 [见 (18)]

$$\frac{R^2}{1-R^2}\cdot\frac{N-p}{p-1}=\frac{\hat{\boldsymbol{\beta}}'\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}}{a_{11\cdot2}}\cdot\frac{N-p}{p-1}\tag{32}$$

是检验假设 X_1 在 $x_2,\cdots,x_p$ 上的回归为 0 的普通 F 统计量. 在本书里, 我们首先考虑将多重相关系数看作一个随机变量和一个随机变量向量的关联性度量. 我们将不考虑单变量回归问题. 在第 8 章中, 我们研究当响应变量为向量时的回归.

调整多重相关系数

表达式 (17) 是拟合回归的离差平方和 $a_{11\cdot2}$ 与均值的离差平方和 a_{11} 的比率. 为了得到 σ_{11} 的无偏估计, 当 $\boldsymbol{\beta}=\mathbf{0}$, 我们可以将这些量除以他们的自由度, 分别为 $N-p$ 和 $N-1$. 因此, 我们可以定义一个**调整多重相关系数**, 即

$$1-R^{*2}=\frac{a_{11\cdot2}/(N-p)}{a_{11}/(N-1)}=\frac{N-1}{N-p}(1-R^2),\tag{33}$$

它等价于

$$R^{*2}=R^2-\frac{p-1}{N-p}(1-R^2).\tag{34}$$

这个量比 R^2 小 (除非 $p=1$ 或 $R^2=1$). 它的一个可能的优点是考虑到了 p, 其思想是 p 相对于 N 越大, R^2 偶然很大的可能性就越大.

4.4.3 当总体多重相关系数非零时样本多重相关系数的分布

本小节我们将求当原假设 $\bar{R}=0$ 不真时, R 的分布. 我们将发现这个分布只依赖总体多重相关系数 $\bar{R}$.

首先, 我们来考虑给定 $\boldsymbol{Z}_\alpha^{(2)}=\boldsymbol{z}_\alpha^{(2)}$ $(\alpha=1,\cdots,n)$ 时 $R^2/(1-R^2)=\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}/a_{11\cdot2}$ 的条件分布. 在这些条件下, $Z_{11},\cdots,Z_{1n}$ 是独立分布的, $Z_{1\alpha}$ 服从分布 $N(\boldsymbol{\beta}'z_\alpha^{(2)},\sigma_{11\cdot2})$, 其中 $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}$ 且 $\sigma_{11\cdot2}=\sigma_{11}-\boldsymbol{\sigma}_{(1)}'\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{(1)}$. 取 $\boldsymbol{Y}_\alpha=Z_{1\alpha},\boldsymbol{\Gamma}=\boldsymbol{\beta}'$, $\boldsymbol{w}_\alpha=\boldsymbol{z}_\alpha^{(2)},r=p-1,\boldsymbol{\Phi}=\sigma_{11\cdot2},m=n$, 这些条件就是定理 4.3.3 的条件. 则 $a_{11\cdot2}=a_{11}-\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}$ 对应于 $\sum_{\alpha=1}^m\boldsymbol{Y}_\alpha\boldsymbol{Y}_\alpha'-\boldsymbol{GHG}'$, 且 $a_{11\cdot2}/\sigma_{11\cdot2}$ 服从自由度为 $n-(p-1)$ 的 χ^2 分布. $\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}=(\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)})'\boldsymbol{A}_{22}(\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)})$ 相应于 $\boldsymbol{GHG}'$ 且与 $\sum_\alpha U_\alpha^2$ $(\alpha=n-(n-p)+1,\cdots,n)$ 同分布, 其中 $\mathrm{Var}(U_\alpha)=\sigma_{11\cdot2}$ 且

$$\mathrm{E}(U_{n-p+2},\cdots,U_n)=\boldsymbol{\Gamma F}^{-1}, \tag{35}$$

其中 $\boldsymbol{FHF}'=\boldsymbol{I}$ $[\boldsymbol{H}=\boldsymbol{F}^{-1}(\boldsymbol{F}')^{-1}]$. 则 $\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot2}$ 与 $\sum_\alpha(U_\alpha/\sqrt{\sigma_{11\cdot2}})^2$ 同分布, 其中 $\mathrm{Var}(U_\alpha/\sqrt{\sigma_{11\cdot2}})=1$ 且

$$\begin{aligned}\sum_{\alpha=n-p+2}^{n}\left(\frac{\mathrm{E}(U_\alpha)}{\sqrt{\sigma_{11\cdot2}}}\right)^2&=\frac{1}{\sigma_{11\cdot2}}\boldsymbol{\Gamma F}^{-1}(\boldsymbol{\Gamma F}^{-1})'=\frac{\boldsymbol{\Gamma H\Gamma}'}{\sigma_{11\cdot2}}\\&=\frac{\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}}{\sigma_{11\cdot2}}.\end{aligned} \tag{36}$$

因此 (条件地) $\boldsymbol{a}_{(1)}'\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}/\sigma_{11\cdot2}$ 服从自由度为 $p-1$, 非中心化参数为 $\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot2}$ 的非中心 χ^2 分布. (见定理 5.4.1.) 我们可以得到下面的定理.

定理 4.4.4 *设 R 是基于 N 个观测 $(x_{11},\boldsymbol{x}_1^{(2)}),\cdots,(x_{1N},\boldsymbol{x}_N^{(2)})$ 的 $X_{(1)}$ 和 $\boldsymbol{X}^{(2)\prime}=(X_2,\cdots,X_p)$ 的样本多重相关系数. 给定 $\boldsymbol{x}_\alpha^{(2)}$, $[R^2/(1-R^2)][(N-p)/(p-1)]$ 的条件分布是自由度为 $p-1$ 和 $N-p$, 非中心化参数为 $\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot2}$ 的非中心 F 分布.*

$F=[R^2/(1-R^2)][(N-p)/(p-1)]$ 的条件密度 (由定理 5.4.1) 是

$$\begin{aligned}&\frac{(p-1)\exp\left[-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot2}\right]}{(N-p)\Gamma[\frac{1}{2}(N-p)]}\\&\cdot\sum_{\alpha=0}^{\infty}\frac{\left(\frac{\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}}{2\sigma_{11\cdot2}}\right)^\alpha\left[\frac{(p-1)f}{N-p}\right]^{\frac{1}{2}(p-1)+\alpha-1}\Gamma[\frac{1}{2}(N-1)+\alpha]}{\alpha!\Gamma[\frac{1}{2}(p-1)+\alpha]\left[1+\frac{(p-1)f}{N-p}\right]^{\frac{1}{2}(N-1)+\alpha}},\end{aligned} \tag{37}$$

且 $W=R^2$ 的条件密度是 $(\mathrm{d}f=[(N-p)/(p-1)](1-w)^{-2}\mathrm{d}w)$

$$\begin{aligned}&\frac{\exp\left[-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot2}\right]}{\Gamma[\frac{1}{2}(N-p)]}(1-w)^{\frac{1}{2}(N-p)-1}\\&\cdot\sum_{\alpha=0}^{\infty}\frac{\left(\frac{\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}}{2\sigma_{11\cdot2}}\right)^\alpha w^{\frac{1}{2}(p-1)+\alpha-1}\Gamma[\frac{1}{2}(N-1)+\alpha]}{\alpha!\Gamma[\frac{1}{2}(p-1)+\alpha]}.\end{aligned} \tag{38}$$

为了得到非条件密度, 我们需要用 $\boldsymbol{Z}_1^{(2)},\cdots,\boldsymbol{Z}_n^{(2)}$ 的密度乘以 (38) 来获得 W 和 $\boldsymbol{Z}_1^{(2)},\cdots,\boldsymbol{Z}_n^{(2)}$ 的联合密度, 继而对后者积分得到 W 的边缘密度. 我们有

$$\begin{aligned}\frac{\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}}{\sigma_{11\cdot 2}}&=\frac{\boldsymbol{\beta}'\sum_{\alpha=1}^{n}\boldsymbol{z}_\alpha^{(2)}\boldsymbol{z}_\alpha^{(2)\prime}\boldsymbol{\beta}}{\sigma_{11\cdot 2}}\\&=\sum_{\alpha=1}^{n}\left(\frac{\boldsymbol{\beta}'\boldsymbol{z}_\alpha^{(2)}}{\sqrt{\sigma_{11\cdot 2}}}\right)^2.\end{aligned}\tag{39}$$

因为 $\boldsymbol{Z}_\alpha^{(2)}$ 的分布是 $N(\mathbf{0},\boldsymbol{\Sigma}_{22})$, 则 $\boldsymbol{\beta}'\boldsymbol{Z}_\alpha^{(2)}/\sqrt{\sigma_{11\cdot 2}}$ 的分布是正态的, 其均值为零, 方差为

$$\begin{aligned}\mathrm{E}\left(\frac{\boldsymbol{\beta}'\boldsymbol{Z}_\alpha^{(2)}}{\sqrt{\boldsymbol{\Sigma}_{11\cdot 2}}}\right)^2&=\frac{\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{Z}_\alpha^{(2)}\boldsymbol{Z}_\alpha^{(2)\prime}\boldsymbol{\beta})}{\sigma_{11\cdot 2}}\\&=\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}{\sigma_{11}-\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}}=\frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}/\sigma_{11}}{1-\boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}/\sigma_{11}}\\&=\frac{\bar{R}^2}{1-\bar{R}^2}.\end{aligned}\tag{40}$$

因此 $(\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot 2})/[\bar{R}^2/(1-\bar{R}^2)]$ 服从自由度为 n 的 χ^2 分布. 设 $\bar{R}^2/(1-\bar{R}^2)=\phi$. 则 $\boldsymbol{\beta}'\boldsymbol{A}_{22}\boldsymbol{\beta}/\sigma_{11\cdot 2}=\phi\chi_n^2$. 我们计算

$$\begin{aligned}&\mathrm{E}\left[\mathrm{e}^{-\frac{1}{2}\phi\chi_n^2}\left(\frac{\phi\chi_n^2}{2}\right)^\alpha\right]\\=&\frac{\phi^\alpha}{2^\alpha}\int_0^\infty u^\alpha\mathrm{e}^{-\frac{1}{2}\phi u}\frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}u^{\frac{1}{2}n-1}\mathrm{e}^{-\frac{1}{2}u}\mathrm{d}u\\=&\frac{\phi^\alpha}{2^\alpha}\int_0^\infty\frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}u^{\frac{1}{2}n+\alpha-1}\mathrm{e}^{-\frac{1}{2}(1+\phi)u}\mathrm{d}u\\=&\frac{\phi^\alpha}{(1+\phi)^{\frac{1}{2}n+\alpha}}\frac{\Gamma(\frac{1}{2}n+\alpha)}{\Gamma(\frac{1}{2}n)}\int_0^\infty\frac{1}{2^{\frac{1}{2}n+\alpha}\Gamma(\frac{1}{2}n+\alpha)}v^{\frac{1}{2}n+\alpha-1}\mathrm{e}^{-\frac{1}{2}v}\mathrm{d}v\\=&\frac{\phi^\alpha}{(1+\phi)^{\frac{1}{2}n+\alpha}}\frac{\Gamma(\frac{1}{2}n+\alpha)}{\Gamma(\frac{1}{2}n)}.\end{aligned}\tag{41}$$

将这个结果应用到 (38), 我们得到 R^2 的密度

$$\frac{(1-R^2)^{\frac{1}{2}(n-p-1)}(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma[\frac{1}{2}(n-p+1)]\Gamma(\frac{1}{2}n)}\sum_{\mu=0}^{\infty}\frac{(\bar{R}^2)^\mu(R^2)^{\frac{1}{2}(p-1)+\mu-1}\Gamma^2(\frac{1}{2}n+\mu)}{\mu!\Gamma[\frac{1}{2}(p-1)+\mu]}.\tag{42}$$

Fisher (1928) 求出了这个分布. 它也可以写为

$$\begin{aligned}&\frac{\Gamma(\frac{1}{2}n)(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma[\frac{1}{2}(n-p+1)]\Gamma[\frac{1}{2}(p-1)]}(R^2)^{\frac{1}{2}(p-3)}(1-R^2)^{\frac{1}{2}(n-p-1)}\\&\cdot F\left[\frac{1}{2}n,\frac{1}{2}n;\frac{1}{2}(p-1);R^2\bar{R}^2\right],\end{aligned}\tag{43}$$

其中, F 是 4.2 节的 (41) 定义的超几何函数.

当 $n-p+1$ 为偶数时, 这个密度可有另外的形式. 我们有

$$\begin{aligned}
&\sum_{\mu=0}^{\infty}\frac{(R^2\bar{R}^2)^{\mu}}{\mu!}\frac{\Gamma^2(\frac{1}{2}n+\mu)}{\Gamma[\frac{1}{2}(p-1)+\mu]} \\
&= \sum_{\mu=0}^{\infty}\frac{(R^2\bar{R}^2)^{\mu}}{\mu!}\Gamma(\frac{1}{2}n+\mu)\left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)}t^{\frac{1}{2}n+\mu-1}\Bigg|_{t=1} \\
&= \left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)}t^{\frac{1}{2}n-1}\sum_{\mu=0}^{\infty}\frac{(t\bar{R}^2R^2)^{\mu}}{\mu!}\frac{\Gamma(\frac{1}{2}n+\mu)}{\Gamma(\frac{1}{2}n)}\Bigg|_{t=1}\Gamma(\frac{1}{2}n) \\
&= \Gamma\left(\frac{1}{2}n\right)\left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)}t^{\frac{1}{2}n-1}(1-tR^2\bar{R}^2)^{-\frac{1}{2}n}\Bigg|_{t=1}.
\end{aligned}\tag{44}$$

因此这个密度是

$$\begin{aligned}
&\frac{(1-\bar{R}^2)^{\frac{1}{2}n}(R^2)^{\frac{1}{2}(p-3)}(1-R^2)^{\frac{1}{2}(n-p-1)}}{\Gamma[\frac{1}{2}(n-p+1)]} \\
&\cdot\left(\frac{\partial}{\partial t}\right)^{\frac{1}{2}(n-p+1)}t^{\frac{1}{2}n-1}(1-tR^2\bar{R}^2)^{-\frac{1}{2}n}\Bigg|_{t=1}.
\end{aligned}\tag{45}$$

定理 4.4.5 *基于容量为 $N=n+1$ 的样本, X_1 和 $(X_2,\cdots,X_p)$ 的样本多重相关系数的平方 R^2 的密度是 (42) 或 (43) [或当 $n-p-1$ 是偶数时, 为 (45)], 其中 $\bar{R}^2$ 是相应的总体多重相关系数.*

R 的矩是

$$\begin{aligned}
\mathrm{E}(R^h)&=\frac{(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma[\frac{1}{2}(n-p+1)]\Gamma(\frac{1}{2}n)}\sum_{\mu=0}^{\infty}\frac{(\bar{R}^2)^{\mu}\Gamma^2(\frac{1}{2}n+\mu)}{\Gamma[\frac{1}{2}(p-1)+\mu]\mu!} \\
&\quad\cdot\int_0^1(1-R^2)^{\frac{1}{2}(n-p+1)-1}(R^2)^{\frac{1}{2}(p+h-1)+\mu-1}\mathrm{d}(R^2) \\
&=\frac{(1-\bar{R}^2)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}\sum_{\mu=0}^{\infty}\frac{(\bar{R}^2)^{\mu}\Gamma^2(\frac{1}{2}n+\mu)\Gamma[\frac{1}{2}(p+h-1)+\mu]}{\mu!\Gamma[\frac{1}{2}(p-1)+\mu]\Gamma[\frac{1}{2}(n+h)+\mu]}.
\end{aligned}\tag{46}$$

样本多重相关系数趋向过分估计总体多重相关系数. 样本多重相关系数是 x_1 和 $\boldsymbol{x}^{(2)}$ 的线性组合的极大样本相关系数, 因此大于 x_1 与 $\boldsymbol{\beta}'\boldsymbol{x}^{(2)}$ 的样本相关系数. 但是后者是相应于 x_1 与 $\boldsymbol{\beta}'\boldsymbol{x}^{(2)}$ 之间的简单总体相关系数的简单样本相关系数, 这里的简单总体相关系数就是总体多重相关系数 $\bar{R}$.

假设 R_1 是两样本中第一个的多重相关系数且 $\hat{\boldsymbol{\beta}}_1$ 是 $\boldsymbol{\beta}$ 的估计, 则在第二个样本中, x_1 和 $\hat{\boldsymbol{\beta}}_1'\boldsymbol{x}^{(2)}$ 的简单相关系数将趋于小于 R_1, 特别地, 它小于第二个样本的多重相关系数 R_2. 这个现象称为"多重相关系数的收缩".

Kramer (1963) 和 Lee (1972) 给出了关于 R 的上分位数表. Gajjar (1967),

Gurland (1968), Gurland and Milton (1970), Khatri (1966) 和 Lee (1917b) 给出了 $R^2/(1-R^2)$ 的近似分布并得到大样本结果.

4.4.4 多重相关系数检验的一些优良性质

定理 4.4.6 给定来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 给定某显著性水平, 所有基于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$ 的关于变换

$$\begin{aligned} &\bar{x}_1^*=c\bar{x}_1+d, \qquad \bar{\boldsymbol{x}}^{(2)*}=\boldsymbol{C}\bar{\boldsymbol{x}}^{(2)}+\boldsymbol{d}, \\ &a_{11}^*=c^2a_{11}, \qquad \boldsymbol{a}_{(1)}^*=c\boldsymbol{C}\boldsymbol{a}_{(1)}, \qquad \boldsymbol{A}_{22}^*=\boldsymbol{C}\boldsymbol{A}_{22}\boldsymbol{C}', \end{aligned} \tag{47}$$

不变的 $\bar{R}=0$ 的检验中, 任何由 R 大于某常数定义的临界区域是一致最优的.

证明 多重相关系数 R 在这些变换下是不变的, 且充分统计量的任何变换不变函数都是 R 的函数. (见习题 4.34.) 因此, 任何不变检验必基于 R. 把奈曼–皮尔逊基本引理应用到检验原假设 $\bar{R}=0$ 及其备择 $\bar{R}=\bar{R}_0>0$, 我们得到, 在指定显著性水平上的最优检验是基于下面两个密度的比, 那就是 $\bar{R}=\bar{R}_0$ 时 R 的密度, 即 (42) 乘以 $2R$ [因为 (42) 是 R^2 的密度], 与 $R=0$ 时的密度, 即 (22). 这个比是一个正常数乘以

$$\sum_{\mu=0}^{\infty}\frac{(\bar{R}_0^2)^\mu\Gamma^2(\frac{1}{2}n+\mu)}{\mu!\Gamma[\frac{1}{2}(p-1)+\mu]}R^{p-2+2\mu}. \tag{48}$$

因为当 $R\geqslant 0$ 时, (48) 是 R 的一个增函数, 所以满足 (48) 大于一个常数的 R 集是一个 R 大于一个常数的区间. ■

定理 4.4.7 基于 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 给定某显著性水平时, 在所有 $\bar{R}=0$ 的功效只依赖 $\bar{R}$ 的检验中, 监界区域是 R 大于一个常数的检验是一致最优的.

定理 4.4.7 由定理 4.4.6 可得, 类似于由定理 5.6.1 得到定理 5.6.4.

4.5 椭球等高分布

4.5.1 椭球等高观测

假设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是 N 个来自 p 维随机向量 $\boldsymbol{X}$ 的独立观测, $\boldsymbol{X}$ 的密度是

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})]. \tag{1}$$

样本协方差阵 $\boldsymbol{S}$ 是协方差阵 $\boldsymbol{\Sigma}=[\mathrm{E}(R^2/p)]\boldsymbol{\Lambda}$ 的一个无偏估计, 其中 $R^2=(\boldsymbol{X}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\nu})$ 且 $\mathrm{E}(R^2)<\infty$. $\rho_{ij}=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}=\lambda_{ij}/\sqrt{\lambda_{ii}\lambda_{jj}}$ 的一个估计是 $r_{ij}=s_{ij}/\sqrt{s_{ii}s_{jj}}, i,j=1,\cdots,p$. r_{ij} 的小样本分布通常比较难求得, 但是可由 3.6 节的 (13) 给出的 $\sqrt{N}(\boldsymbol{S}-\boldsymbol{\Sigma})$ 的极限正态分布得到其渐近分布.

首先, 我们用定理 4.2.3 和定理 3.6.5 证明一个关于样本协方差阵 $\boldsymbol{S}$ 的函数的渐近分布. 定义

$$\boldsymbol{s} = \mathrm{vec}\boldsymbol{S}, \quad \boldsymbol{\sigma} = \mathrm{vec}\boldsymbol{\Sigma}. \tag{2}$$

定理 4.5.1 设 $\boldsymbol{f}(\boldsymbol{s})$ 是一个向量值函数, 它的每个分量在 $\boldsymbol{s}=\boldsymbol{\sigma}$ 有非零微分. 假设 $\boldsymbol{S}$ 是来自 (1) 的样本的协方差阵, 满足 $\mathrm{E}(R^4) < \infty$. 则

$$\begin{aligned}\sqrt{N}[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] &= \frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\sqrt{N}(\boldsymbol{s}-\boldsymbol{\sigma}) + o_p(1) \\ &\xrightarrow{d} N\left\{\mathbf{0}, \frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}[2(1+\kappa)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \kappa\boldsymbol{\sigma}\boldsymbol{\sigma}']\left(\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right)'\right\}.\end{aligned} \tag{3}$$

推论 4.5.1 如果对所有 $c>0$ 和所有正定矩阵 $\boldsymbol{S}$ 以及使定理 4.5.1 成立的条件, 有

$$\boldsymbol{f}(c\boldsymbol{s}) = \boldsymbol{f}(\boldsymbol{s}), \tag{4}$$

则

$$\sqrt{N}[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] \xrightarrow{d} N\left[\mathbf{0}, 2(1+\kappa)\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\left(\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right)'\right]. \tag{5}$$

证明 由 (4), 我们推得

$$\mathbf{0} = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial c} = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial \boldsymbol{s}'}\frac{\partial (c\boldsymbol{s})}{\partial c} = \frac{\partial \boldsymbol{f}(c\boldsymbol{s})}{\partial \boldsymbol{s}'}\boldsymbol{s}. \tag{6}$$

即

$$\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\boldsymbol{\sigma} = \mathbf{0}. \quad ■ \tag{7}$$

推论 4.5.1 的结论可以写成

$$\frac{\sqrt{N}}{\sqrt{1+\kappa}}[\boldsymbol{f}(\boldsymbol{s}) - \boldsymbol{f}(\boldsymbol{\sigma})] \xrightarrow{d} N\left[\mathbf{0}, 2\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\left(\frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}\right)'\right]. \tag{8}$$

特别地, 当样本来自正态分布时, (8) 中极限正态分布成立. 用 κ 的一致估计 $\hat{\kappa}$ 代替 κ, 推论也是成立的. 例如, 3.6 节的 (16) 给出了 $1+\hat{\kappa}$ 的一个一致估计, 是

$$1+\hat{\kappa} = \sum_{\alpha=1}^{N}\left[(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})\right]^2/[Np(p+2)]. \tag{9}$$

一个样本相关系数 (比如 $f(\boldsymbol{s}) = r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$) 或一个这样的相关系数的集合是 $\boldsymbol{S}$ 的一个函数, 满足在比例变换下不变, 即满足 (4).

推论 4.5.2 在定理 4.5.1 的条件下,

$$\sqrt{\frac{N}{1+\hat{\kappa}}}\frac{(r_{ij}-\rho_{ij})}{\sqrt{1-r_{ij}^2}} \xrightarrow{d} N(0,1). \tag{10}$$

当观测是正态分布时,

$$\sqrt{\frac{N}{1+\hat{\kappa}}}\left(\frac{1}{2}\ln\frac{1+r_{ij}}{1-r_{ij}} - \frac{1}{2}\ln\frac{1+\rho_{ij}}{1-\rho_{ij}}\right) \xrightarrow{d} N(0,1). \tag{11}$$

当然, (11) 对 (10) 的任何改进依赖样本的分布.

偏相关系数, 例如 $r_{ij\cdot q+1,\cdots,p}, i,j=1,\cdots,q$, 也是 $\boldsymbol{S}$ 的不变函数.

推论 4.5.3　*在定理 4.5.1 的条件下,*

$$\sqrt{\frac{N}{1+\hat{\kappa}}}(r_{ij\cdot q+1,\cdots,p}-\rho_{ij\cdot q+1,\cdots,p}) \xrightarrow{d} N(0,1). \tag{12}$$

现在我们考虑当总体的多重相关系数平方 $\bar{R}^2$ 为 0 时, 多重相关系数平方 R^2 的渐近分布. 我们用 4.4 节的符号. $\bar{R}^2=0$ 等价于 $\boldsymbol{\sigma}_{(1)}=\boldsymbol{0}$. 由于 X_1 和 $\boldsymbol{X}_2^{(2)}=(X_2,\cdots,X_p)'$ 的样本和总体多重相关系数是关于 4.4 节的线性变换 (47) 不变的, 从而为了研究 R^2 的分布, 我们可以假定 $\boldsymbol{\mu}=\boldsymbol{0}$ 和 $\boldsymbol{\Sigma}=\boldsymbol{I}_p$. 在这种情况下, $s_{11}\xrightarrow{P}1$, $\boldsymbol{s}_{(1)}\xrightarrow{P}\boldsymbol{0}$ 和 $\boldsymbol{S}_{22}\xrightarrow{P}\boldsymbol{I}_{p-1}$. 从而, 对于 $k,i\neq 1$ 和 $j=l=1$, 由引理 3.6.1 得到

$$\mathrm{E}(\boldsymbol{s}_{(1)}\boldsymbol{s}_{(1)}')=\left(\frac{1}{n}+\frac{\kappa}{N}\right)\boldsymbol{I}_{p-1}. \tag{13}$$

定理 4.5.2　*在定理 4.5.1 的条件下,*

$$\sqrt{\frac{N}{1+\kappa}}\boldsymbol{s}_{(1)} \xrightarrow{d} N(\boldsymbol{0},\boldsymbol{I}_{p-1}). \tag{14}$$

推论 4.5.4　*在定理 4.5.1 的条件下,*

$$\frac{NR^2}{1+\hat{\kappa}}=\frac{N\boldsymbol{s}_{(1)}'\boldsymbol{S}_{22}^{-1}\boldsymbol{s}_{(1)}}{(1+\hat{\kappa})s_{11}} \xrightarrow{d} \chi_{p-1}^2. \tag{15}$$

4.5.2　椭球等高矩阵分布

现在我们考虑基于向量球面模型 $g(\mathrm{tr}\boldsymbol{Y}'\boldsymbol{Y})$ 的模型

$$|\boldsymbol{\Lambda}|^{-N/2}g[\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')']. \tag{16}$$

$\boldsymbol{\nu}$ 和 $\boldsymbol{\Sigma}=(\mathrm{E}(R^2/p))\boldsymbol{\Lambda}$ 的无偏估计是 $\bar{\boldsymbol{x}}=(1/N)\boldsymbol{X}'\boldsymbol{\varepsilon}_N$ 和 $\boldsymbol{S}=(1/n)\boldsymbol{A}$, 其中 $\boldsymbol{A}=(\boldsymbol{X}-\boldsymbol{\varepsilon}_n\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')$.

因为

$$(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')=\boldsymbol{A}+N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})', \tag{17}$$

$\boldsymbol{A}$ 和 $\bar{\boldsymbol{x}}$ 是一个完全充分统计量集合.

$\boldsymbol{\nu}$ 和 $\boldsymbol{\Lambda}$ 的极大似然估计是 $\hat{\boldsymbol{\nu}}=\bar{\boldsymbol{x}}$ 和 $\hat{\boldsymbol{\Lambda}}=(p/w_g)\boldsymbol{A}$. $\rho_{ij}=\lambda_{ij}/\sqrt{\lambda_{ii}\lambda_{jj}}=\sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ 的极大似然估计是 $\hat{\rho}_{ij}=a_{ij}/\sqrt{a_{ii}a_{jj}}=s_{ij}/\sqrt{s_{ii}s_{jj}}$ (定理 3.6.4).

样本相关系数 r_{ij} 是 $f(\boldsymbol{X})$ 的一个函数, 它满足定理 3.6.5 中的 (45) 和 (46), 从而在 $g[\mathrm{tr}(\cdot)]$ 为任意密度时和 $g[\mathrm{tr}(\cdot)]=\mathrm{const}\ \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\cdot)}$ 为正态密度时的分布相同. 类似地, 偏相关系数 $r_{ij\cdot q+1,\cdots,p}$ 和多重相关系数 R^2 也满足那些条件, 从而结论成立.

定理 4.5.3　*当 $\boldsymbol{X}$ 有向量椭球密度 (16) 时, r_{ii}, $r_{ij\cdot q+1,\cdots,p}$ 和 R^2 的分布是观测为正态分布时的分布.*

由定理 4.5.3 得, r_{ij}, $r_{ij\cdot q+1,\cdots,p}$ 和 R^2 的渐近分布与来自正态分布的抽样一样.

有密度的左球矩阵 $\boldsymbol{Y}$ 族是 $g(\boldsymbol{Y}'\boldsymbol{Y})$ 族. 设 $\boldsymbol{X}=\boldsymbol{Y}\boldsymbol{C}'+\boldsymbol{\varepsilon}_N\boldsymbol{\nu}'$, 其中 $\boldsymbol{C}'\boldsymbol{\Lambda}^{-1}\boldsymbol{C}=\boldsymbol{I}$, 即 $\boldsymbol{\Lambda}=\boldsymbol{C}\boldsymbol{C}'$. 则 $\boldsymbol{X}$ 有密度

$$|\boldsymbol{C}|^{-N}g\left[\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')(\boldsymbol{C}')^{-1}\right]. \tag{18}$$

我们现在求矩阵 $\boldsymbol{Y}$ 的一个随机表示.

引理 4.5.1 设 $\boldsymbol{V}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p)$, 其中 $\boldsymbol{v}_i$ 是一个 N 维向量, $i=1,\cdots,p$. 循环定义 $\boldsymbol{w}_1=\boldsymbol{v}_1$,

$$\boldsymbol{w}_i=\boldsymbol{v}_i-\sum_{j=1}^{i-1}\frac{\boldsymbol{v}_i'\boldsymbol{w}_j}{\boldsymbol{w}_j'\boldsymbol{w}_j}\boldsymbol{w}_j,\quad i=2,\cdots,p. \tag{19}$$

设 $\boldsymbol{u}_i=\boldsymbol{w}_i/\|\boldsymbol{w}_i\|$. 则 $\|\boldsymbol{u}_i\|=1, i=1,\cdots,p$, $\boldsymbol{u}_i'\boldsymbol{u}_j=0, i\neq j$. 进而,

$$\boldsymbol{V}=\boldsymbol{U}\boldsymbol{T}', \tag{20}$$

其中 $\boldsymbol{U}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_p)$; $t_{ii}=\|\boldsymbol{w}_i\|, i=1,\cdots,p$; $t_{ij}=\boldsymbol{v}_i'\boldsymbol{w}_j/\|\boldsymbol{w}_j\|=\boldsymbol{v}_i'\boldsymbol{u}_j, j=1,\cdots,i-1, i=1,\cdots,p$; $t_{ij}=0, i<j$.

引理的证明在 7.2 节的第一部分给出, 它和附录 (A.5.1 节) 中的格拉姆–施密特正交化一样. 这个引理推广了 3.2 节的构造, 见图 3.1 和图 7.1.

注意到 $\boldsymbol{T}$ 是下三角矩阵, $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_p$ 和 $\boldsymbol{V}'\boldsymbol{V}=\boldsymbol{T}\boldsymbol{T}'$. 最后一个等式, $t_{ii}\geqslant 0$ $(i=1,\cdots,p)$ 和 $t_{ij}=0$ $(i<j)$, 可以唯一解出 $\boldsymbol{T}$. 因此, $\boldsymbol{T}$ 是 $\boldsymbol{V}'\boldsymbol{V}$ (及其限制) 的一个函数.

设 $\boldsymbol{Y}(N\times p)$ 有密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$, 且设 $\boldsymbol{O}_N$ 是一个 $N\times N$ 正交矩阵. 则 $\boldsymbol{Y}^*=\boldsymbol{O}_N\boldsymbol{Y}$ 有密度 $g(\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*)$. 因此, $\boldsymbol{Y}^*=\boldsymbol{O}_N\boldsymbol{Y}\overset{d}{=}\boldsymbol{Y}$. 设 $\boldsymbol{Y}^*=\boldsymbol{U}^*\boldsymbol{T}^{*\prime}$, 其中 $t_{ii}^*>0$ $(i=1,\cdots,p)$ 和 $t_{ij}^*=0$ $(i<j)$. 由 $\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*=\boldsymbol{Y}'\boldsymbol{Y}$ 知, $\boldsymbol{T}^*\boldsymbol{T}^{*\prime}=\boldsymbol{T}\boldsymbol{T}'$, 从而 $\boldsymbol{T}^*=\boldsymbol{T}, \boldsymbol{Y}^*=\boldsymbol{U}^*\boldsymbol{T}$, 且 $\boldsymbol{U}^*=\boldsymbol{O}_N\boldsymbol{U}\overset{d}{=}\boldsymbol{U}$. 设满足 $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_p$ 的 $\boldsymbol{U}$ $(N\times p)$ 的空间表示为 $O(N\times p)$.

定义 4.5.1 如果 $\boldsymbol{U}(N\times p)$ 满足 $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_p$ 和 $\boldsymbol{O}_N\boldsymbol{U}\overset{d}{=}\boldsymbol{U}$, 对所有正交 $\boldsymbol{O}_N$, 则 $\boldsymbol{U}$在 $O(N\times p)$ 上均匀分布.

满足 $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_p$ 的 $\boldsymbol{U}$ 的空间称为斯蒂芬 (Steifel) 流形. 定义 4.5.1 的概率测度称为哈尔 (Haar) 不变分布. 对所有正交的 $\boldsymbol{O}_N$, $\boldsymbol{O}_N\boldsymbol{U}\overset{d}{=}\boldsymbol{U}$ 的性质唯一定义了此 (正态化) 度量.

定理 4.5.4 如果 $\boldsymbol{Y}$ $(N\times p)$ 有密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$, 则由 $\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{T}', \boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}_p, t_{ii}>0$ $(i=1,\cdots,p)$ 和 $t_{ij}=0$ $(i<j)$ 定义的 $\boldsymbol{U}$ 在 $O(N\times p)$ 上均匀分布.

推论 7.2.1 的证明说明了对任意 $g(\cdot)$, $\boldsymbol{T}$ 的密度是

$$\prod_{i=1}^{p}\left\{C\left[\frac{1}{2}(N+1-i)\right]t_{ii}^{N-i}\right\}g(\mathrm{tr}\boldsymbol{T}\boldsymbol{T}'), \tag{21}$$

其中, $C(\cdot)$ 由 2.7 节的 (8) 定义.

密度为 $g(\boldsymbol{Y}'\boldsymbol{Y})$ 的 $\boldsymbol{Y}$ $(N\times p)$ 的随机表示是

$$\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{T}', \tag{22}$$

其中 $\boldsymbol{U}$ $(N\times p)$ 在 $O(N\times p)$ 上均匀分布, $\boldsymbol{T}$ 是对角线元素为正数的下三角矩阵并且有密度 (21).

定理 4.5.5 *设 $\boldsymbol{f}(\boldsymbol{X})$ 是一个 $\boldsymbol{X}$ $(N\times p)$ 的向量值函数, 满足对所有的 $\boldsymbol{\nu}$ 有*

$$\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')=\boldsymbol{f}(\boldsymbol{X}) \tag{23}$$

和对所有的 $\boldsymbol{G}(p\times p)$ 有

$$\boldsymbol{f}(\boldsymbol{X}\boldsymbol{G}')=\boldsymbol{f}(\boldsymbol{X}) \tag{24}$$

则当 $\boldsymbol{X}$ 有任意密度 (18) 时 $\boldsymbol{f}(\boldsymbol{X})$ 的分布与当 $\boldsymbol{X}$ 有正态密度时 $f(X)$ 的分布一样.

证明 由 (23) 我们得到 $\boldsymbol{f}(\boldsymbol{X})=\boldsymbol{f}(\boldsymbol{Y}\boldsymbol{C}')$, 由 (24) 我们得到 $\boldsymbol{f}(\boldsymbol{Y}\boldsymbol{C}')=\boldsymbol{f}(\boldsymbol{U}\boldsymbol{T}'\boldsymbol{C}')=\boldsymbol{f}(\boldsymbol{U})$, 这对于任意密度和正态密度都是一样的. ■

推论 4.5.5 *设 $\boldsymbol{f}(\boldsymbol{X})$ 是一个 $\boldsymbol{X}$ $(N\times p)$ 的向量值函数, $\boldsymbol{X}$ 有密度 (18), 其中 $\boldsymbol{\nu}=\mathbf{0}$. 假设 (24) 对所有 $\boldsymbol{G}$ $(p\times p)$ 成立. 则当 $\boldsymbol{X}$ 为任意密度 (18) 时 $\boldsymbol{f}(\boldsymbol{X})$ 的分布和当 $\boldsymbol{X}$ 有正态密度 (18) 时 $\boldsymbol{f}(\boldsymbol{X})$ 的分布一样.*

推论 4.5.5 中的条件 (24) 是 $\boldsymbol{f}(\boldsymbol{X})$ 关于线性变换 $\boldsymbol{X}\to\boldsymbol{X}\boldsymbol{G}$ 不变的.

密度 (18) 可以写为

$$|\boldsymbol{C}|^{-1}g\left\{\boldsymbol{C}^{-1}[\boldsymbol{A}+N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'](\boldsymbol{C}')^{-1}\right\}, \tag{25}$$

它可以证明 $\boldsymbol{A}$ 和 $\bar{\boldsymbol{x}}$ 是 $\boldsymbol{\Lambda}=\boldsymbol{C}\boldsymbol{C}'$ 和 $\boldsymbol{\nu}$ 的充分统计量的完全集合.

习 题

4.1 (4.2.1 节) 描画

$$k_N(r)\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma(\frac{1}{2}N-1)\sqrt{\pi}}(1-r^2)^{\frac{1}{2}(N-4)}$$

在 **(a)** $N=3$, **(b)** $N=4$, **(c)** $N=5$ 和 **(d)** $N=10$ 时的图形.

4.2 (4.2.1 节) 利用习题 3.1 的数据求假设 X_1 和 X_2 独立及其所有非独立备择的显著性水平为 0.01 的检验.

4.3 (4.2.1 节) 假设容量为 10 的样本的相关系数为 0.65. 求独立性假设及其正相关备择的显著性水平为 0.05 的检验.

4.4 (4.2.1 节) 假设容量为 20 的样本的相关系数为 0.65. 在显著性水平为 0.05 时, 求总体相关系数是 0.4 的假设及总体相关系数大于 0.4 的备择的检验.

4.5 (4.2.1 节) 由 $N=15$ 个观测分别求检验 $\rho=0$ 及以下备择的显著性水平为 0.01 的分位数: **(a)** $\rho\neq 0$, **(b)** $\rho>0$, **(c)** $\rho<0$.

4.6 (4.2.2 节) 由 $N=20$ 个观测求检验 $\rho=0$ 及以下备择的显著性水平为 0.01 的分位数: **(a)** $\rho\neq 0.6$, **(b)** $\rho>0.6$, **(c)** $\rho<0.6$.

4.7 (4.2.2 节) 列出习题 4.5 中的检验在 $\rho=-1(0.2)1$ 的功效函数. 画出每个功效函数的图.

4.8 (4.2.2 节) 列出习题 4.6 中的检验在 $\rho=-1(0.2)1$ 的功效函数. 画出每个功效函数的图.

4.9 (4.2.2 节) 利用习题 3.1 中的数据, 求 ρ_{12} 的置信度为 0.99 的 (双边) 置信区间.

4.10 (4.2.2 节) 假设 $N=10, r=0.795$. 求 ρ 的一个置信度为 0.95 的单边置信区间 (以 $(r_0,\ 1)$ 的形式).

4.11 (4.2.3 节) 利用 Fisher z 求假设 $\rho=0.7$ 及其备择 $\rho\neq 0.7$ 的水平为 0.05 的检验, 取 $r=0.5$ 和 $N=50$.

4.12 (4.2.3 节) 利用 Fisher z 求假设 $\rho_1=\rho_2$ 及其备择 $\rho_1\neq\rho_2$ 的水平为 0.01 的检验, 取 $r_1=0.5, N_1=40, r_2=0.6, N_2=40$.

4.13 (4.2.3 节) 基于样本相关系数 -0.7 $(N=30)$ 和 -0.6 $(N=40)$, 利用 Fisher z 估计 ρ.

4.14 (4.2.3 节) 基于样本相关系数 0.65, 样本容量为 25, 利用 Fisher z 求 ρ 的置信度为 0.95 的置信区间.

4.15 (4.2.2 节) 证明当 $N=2$ 和 $\rho=0$ 时, $\Pr\{r=1\}=\Pr\{r=-1\}=\frac{1}{2}$.

4.16 (4.2 节) 给定 ρ 和 N 的值, 设 $k_N(r,\rho)$ 是样本相关系数 r 的密度. 证明 r 有单调似然比, 即证明如果 $\rho_1>\rho_2$, 则 $k_N(r,\rho_1)/k_N(r,\rho_2)$ 是关于 r 单增的. [提示: 利用 (40), 证明如果

$$F\left[\frac{1}{2},\frac{1}{2};n+\frac{1}{2};\frac{1}{2}(1+\rho r)\right]=\sum_{\alpha=0}^{\infty}c_\alpha(1+\rho r)^\alpha=g(r,\rho)$$

有一个单调比, 则 $k_N(r,\rho)$ 同样如此. 证明

$$\frac{\partial^2}{\partial\rho\partial r}\ln g(r,\rho)=\frac{\sum_{\alpha,\beta=0}^{\infty}c_\alpha c_\beta\left[(\alpha-\beta)^2 r\rho+(\alpha+\beta)\right](1+r\rho)^{\alpha+\beta-2}}{2\left[\sum_{\alpha=0}^{\infty}c_\alpha(1+r\rho)^\alpha\right]^2},$$

如果 $(\partial^2/\partial\rho\partial r)\ln g(r,\rho)>0$, 则 $g(r,\rho)$ 有单调比. 通过证明对每个 α, 关于 β 的和是正的来证明上式的分子是正的, 利用 $c_{\alpha+1}<\frac{1}{2}c_\alpha$.]

4.17 (4.2 节) 证明所有基于 r 的 ρ_0 及其备择 ρ_1 $(>\rho_0)$ 的检验中, 那些拒绝区域形如 $r>c$ 的方法是最优的. [提示: 可由习题 4.16 得.]

4.18 (4.2 节) 证明所有基于 r 的 $\rho=\rho_0$ 及其备择 $\rho>\rho_0$ 的检验中, 拒绝区域形如 $r>c$ 的方法是一致最优的.

4.19 (4.2 节) 通过证明 $h(r)=k_N(r,\rho_1)/k_N(r,\rho_2)$ 当 $\rho_1>\rho_2$ 时是单增的, 证明当 $r>0,\rho>0$ 时 r 有单调似然比. 这里 $h(r)$ 是一个常数乘以 $(\sum_{\alpha=0}^{\infty}c_\alpha\rho_1^\alpha r^\alpha)/(\sum_{\alpha=0}^{\infty}c_\alpha\rho_2^\alpha r^\alpha)$. 证明在 $h'(r)$ 的分子里, r^β 的系数是正的.

4.20 (4.2 节) 证明如果 $\mathbf{\Sigma}$ 是对角的, 则 r_{ij} 和 a_{ii} 的集合是独立分布的. [提示: 利用 r 是比例变换下的不变量以及观测的密度只依赖 a_{ii}.]

4.21 (4.2.1 节) 证明如果 $\rho=0$, 则

$$\mathrm{E}(r^{2m})=\frac{\Gamma[\frac{1}{2}(N-1)]\Gamma(m+\frac{1}{2})}{\sqrt{\pi}\,\Gamma[\frac{1}{2}(N-1)+m]}.$$

4.22　(4.2.2 节)　证明 $f_1(\rho)$ 和 $f_2(\rho)$ 是 ρ 的单调递增函数.

4.23　(4.2.2 节)　证明样本相关系数 r 的密度 [由 (38) 给出] 是

$$\frac{n-1}{\pi}(1-\rho^2)^{\frac{1}{2}n}(1-r^2)^{\frac{1}{2}(n-3)}\int_0^1 \frac{x^{n-1}\mathrm{d}x}{(1-\rho r x)^n\sqrt{1-x^2}}.$$

[提示：用幂级数展开 $(1-\rho r x)^{-n}$, 积分并利用伽玛函数的倍量公式.]

4.24　(4.2 节)　证明 (39) 是 r 的密度. [提示：由习题 2.12, 证明

$$\int_0^\infty\int_0^\infty \mathrm{e}^{-\frac{1}{2}(y^2-2xyz+z^2)}\mathrm{d}y\mathrm{d}z=\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}.$$

则讨论

$$\int_0^\infty\int_0^\infty (yz)^{n-1}\mathrm{e}^{-\frac{1}{2}(y^2-2xyz+z^2)}\mathrm{d}y\mathrm{d}z=\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}.$$

最后证明 (31) 关于 $a_{11}(=y^2)$ 和 $a_{22}(=z^2)$ 的积分是 (39).]

4.25　(4.2 节)　证明 (40) 是 r 的密度. [提示：在 (31) 中, 设 $a_{11}=u\mathrm{e}^{-v}$ 和 $a_{22}=u\mathrm{e}^{v}$, 证明 v $(0\leqslant v<\infty)$ 和 $r(-1\leqslant r\leqslant 1)$ 的密度是

$$\frac{n-1}{\pi\sqrt{2}}(1-\rho^2)^{\frac{1}{2}n}(1-\rho r)^{-n+\frac{1}{2}}(1-r^2)^{\frac{1}{2}(n-3)}v^{-\frac{1}{2}}(1-v)^{n-1}\left[1-\frac{1}{2}(1+\rho r)v\right]^{-\frac{1}{2}}.$$

利用展式

$$(1-y)^{-\frac{1}{2}}=\sum_{j=0}^\infty\frac{\Gamma(j+\frac{1}{2})}{\Gamma(\frac{1}{2})j!}y^j.$$

证明这个积分是 (40).]

4.26　(4.2 节)　证明：对于整数 h,

$$\mathrm{E}(r^{2h+1})=\frac{(1-\rho^2)^{\frac{1}{2}n}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)}\sum_{\beta=0}^\infty\frac{(2\rho)^{2\beta+1}}{(2\beta+1)!}\frac{\Gamma^2\left[\frac{1}{2}(n+1)+\beta\right]\Gamma(h+\beta+\frac{3}{2})}{\Gamma(\frac{1}{2}n+h+\beta+1)},$$

$$\mathrm{E}(r^{2h})=\frac{(1-\rho^2)^{\frac{1}{2}n}}{\sqrt{\pi}\,\Gamma(\frac{1}{2}n)}\sum_{\beta=0}^\infty\frac{(2\rho)^{2\beta}}{(2\beta)!}\frac{\Gamma^2(\frac{1}{2}n+\beta)\Gamma(h+\beta+\frac{1}{2})}{\Gamma(\frac{1}{2}n+h+\beta)}.$$

4.27　(4.2 节)　*t* 分布. 证明：如果 X 和 Y 独立分布, X 服从分布 $N(0,1)$, Y 服从自由度为 m 的 χ^2 分布, 则 $W=X/\sqrt{Y/m}$ 有密度

$$\frac{\Gamma\left[\frac{1}{2}(m+1)\right]}{\sqrt{m}\sqrt{\pi}\,\Gamma(\frac{1}{2}m)}\left(1+\frac{t^2}{m}\right)^{-\frac{1}{2}(m+1)}.$$

[提示：在 X 和 Y 的联合密度中, 设 $x=tw^{\frac{1}{2}}m^{-\frac{1}{2}}$ 并对 w 积分.]

4.28　(4.2 节)　证明

$$\mathrm{E}(r)=\frac{(1-\rho^2)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}\sum_{\beta=0}^\infty\frac{\rho^{2\beta+1}\Gamma^2\left[\frac{1}{2}(n+1)+\beta\right]}{\beta!\Gamma\left[\frac{1}{2}n+\beta+1\right]}.$$

[提示：利用习题 4.26 和伽玛函数的倍量公式.]

4.29　(4.2 节)　证明 $\sqrt{n}(r_{ij}-\rho_{ij}),(i,j)=(1,2),(1,3),(2,3)$, 有联合极限分布, 其方差为 $(1-\rho_{ij}^2)^2$ 且 r_{ij} 和 r_{ik} $(j\neq k)$ 的协方差为 $\frac{1}{2}(2\rho_{jk}-\rho_{ij}\rho_{jk})(1-\rho_{ij}^2-\rho_{ik}^2-\rho_{jk}^2)+\rho_{jk}^2$.

4.30　(4.3.2 节)　基于 $r_{13\cdot 2}=0.097$ 和 $N=20$, 求 $\rho_{13\cdot 2}$ 的置信度为 0.95 的置信区间.

4.31 (4.3.2 节) 取 $r_{12\cdot34} = 0.14$ 和 $N = 40$, 利用 Fisher z 求假设 $\rho_{12\cdot34} = 0$ 及其备择 $\rho_{12\cdot34} \neq 0$ 的显著性水平为 0.01 的检验.

4.32 (4.3 节) 证明不等式 $r_{12\cdot3}^2 \leqslant 1$ 与不等式 $|r_{ij}| \geqslant 0$ 相同, 其中 $|r_{ij}|$ 表示 3×3 的相关系数阵的行列式.

4.33 (4.3 节) *样本偏相关系数的不变性*. 证明 $r_{12\cdot3,\cdots,p}$ 在下列变换下是不变的: $x_{i\alpha}^* = a_i x_{i\alpha} + \boldsymbol{b}_i'\boldsymbol{x}_\alpha^{(3)} + c_i, a_i > 0, i = 1,2, \boldsymbol{x}_\alpha^{(3)*} = \boldsymbol{C}\boldsymbol{x}_\alpha^{(3)} + \boldsymbol{b}, \alpha = 1,\cdots,N$, 其中 $\boldsymbol{x}_\alpha^{(3)} = (x_{3\alpha},\cdots,x_{p\alpha})'$, 并且任何 $\bar{\boldsymbol{x}}$ 和 $\hat{\boldsymbol{\Sigma}}$ 的函数, 如果在这些变换下不变的, 则必是 $r_{12\cdot3,\cdots,p}$ 的函数.

4.34 (4.4 节) *样本多重相关系数的不变性*. 证明 R 是充分统计量 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 的函数, 满足在 $x_{1\alpha}$ 的位移变换和比例变换, 及 $\boldsymbol{x}_\alpha^{(2)}$ 的非奇异线性变换 (即 $x_{1\alpha}^* = cx_{1\alpha} + d, \boldsymbol{x}_\alpha^{(2)*} = \boldsymbol{C}\boldsymbol{x}_\alpha^{(2)} + \boldsymbol{d}, \alpha = 1,\cdots,N$) 下不变, 且每一个 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 的这样的不变函数都是 R 的一个函数.

4.35 (4.4 节) 证明在条件 $Z_{1\alpha} = z_{1\alpha}, \alpha = 1,\cdots,n$ 下, $R^2/(1-R^2)$ 与 $T^2/(N^*-1)$ 同分布, 其中 $T^2 = N^*\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}$ 是基于 $p^* = p-1$ 维向量 $\boldsymbol{X}$ 的 $N^* = n$ 个观测上的, 其均值向量是 $(c/\sigma_{11})\boldsymbol{\sigma}_{(1)}'$ $(nc^2 = \sum z_{1\alpha}^2)$, 协方差阵为 $\boldsymbol{\Sigma}_{22\cdot1} = \boldsymbol{\Sigma}_{22} - (1/\sigma_{11})\boldsymbol{\sigma}_{(1)}\boldsymbol{\sigma}_{(1)}'$. [**提示**: 给定 $Z_{1\alpha} = z_{1\alpha}$ 时 $\boldsymbol{Z}_\alpha^{(2)}$ 的条件分布是 $N[(1/\sigma_{11})\boldsymbol{\sigma}_{(1)}z_{1\alpha}, \boldsymbol{\Sigma}_{22\cdot1}]$. 存在 $n\times n$ 正交矩阵 $\boldsymbol{B}$ 将 $(z_{11},\cdots,z_{1n})$ 变换到 $(c,\cdots,c)$, 并且将 $(Z_{i1},\cdots,Z_{in})$ 变换到 $(Y_{i1},\cdots,Y_{in}), i = 2,\cdots,p$. 设新的 $\boldsymbol{X}_\alpha'$ 为 $(Y_{2\alpha},\cdots,Y_{p\alpha})$.]

4.36 (4.4 节) 证明习题 4.35 中的分布的非中心化参数是 $(a_{11}/\sigma_{11})\bar{R}^2/(1-\bar{R}^2)$.

4.37 (4.4 节) 通过将 a_{11} 的密度乘以习题 4.35 中的密度, 再对 a_{11} 积分, 求 $R^2/(1-R^2)$ 的分布.

4.38 (4.4 节) 证明从 4.2 节的 (38) 得到的 r^2 的密度等价于 4.4 节中 $p = 2$ 时的 (42). [**提示**: 利用伽玛函数的倍量公式.]

4.39 (4.4 节) 证明 (30) 是基于 r 的 $\bar{R} = 0$ 的一致最优检验. [**提示**: 利用奈曼–皮尔逊 (Neyman-Pearson) 基本引理.]

4.40 (4.4 节) 证明 (47) 是基于 R^2 的 $\bar{R}^2$ 的唯一无偏估计.

4.41 习题 3.1 中 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的估计是

$$\bar{\boldsymbol{x}} = (185.72 \quad 151.12 \quad 183.84 \quad 149.24)',$$

$$\boldsymbol{S} = \left(\begin{array}{cc|cc} 95.2933 & 52.8683 & 69.6617 & 46.1117 \\ 52.8683 & 54.3600 & 51.3117 & 35.0533 \\ \hline 69.6617 & 51.3117 & 100.8067 & 56.5400 \\ 46.1117 & 35.0533 & 56.5400 & 45.0233 \end{array}\right).$$

(a) 求给定 (x_1,x_2) 时 (x_3,x_4) 的条件分布的参数估计, 即求 $\boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}$ 和 $\boldsymbol{S}_{22\cdot1} = \boldsymbol{S}_{22} - \boldsymbol{S}_{21}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{12}$.

(b) 求偏相关系数 $r_{34\cdot12}$.

(c) 利用 Fisher z 求 $\rho_{34\cdot12}$ 的置信度为 0.95 的置信区间.

(d) 求 x_3 和 (x_1,x_2) 的样本多重相关系数, 以及 x_4 和 (x_1,x_2) 的样本多重相关系数.

(e) 以显著性水平 0.05 检验假设 x_3 和 (x_1, x_2) 独立以及 x_4 和 (x_1, x_2) 独立.

4.42　设 $\boldsymbol{X}$ 的分量相应计算速度 (X_1), 计算能力 (X_2), 单词记忆力 (X_3), 有用信号记忆力 (X_4) 和无用信号记忆力 (X_5) 的测试得分. 一个容量为 140 的样本观测相关系数是 [Kelley (1928)]

$$\begin{pmatrix} 1.0000 & 0.4248 & 0.0420 & 0.0215 & 0.0573 \\ 0.4248 & 1.0000 & 0.1487 & 0.2489 & 0.2843 \\ 0.0420 & 0.1487 & 1.0000 & 0.6693 & 0.4662 \\ 0.0215 & 0.2489 & 0.6693 & 1.0000 & 0.6915 \\ 0.0573 & 0.2843 & 0.4662 & 0.6915 & 1.0000 \end{pmatrix}.$$

(a) 求给定 X_3 时 X_4 和 X_5 的偏相关系数.

(b) 求给定 X_3, X_4 与 X_5 时, X_1 和 X_2 的偏相关系数.

(c) 求 X_1 和集合 X_3, X_4 及 X_5 的多重相关系数.

(d) 以显著性水平 1% 检验计算速度与其他三个记忆力得分独立的假设.

4.43　(4.3 节)　证明: 如果 $\rho_{ij\cdot q+1,\cdots,p}=0$, 则 $\sqrt{N-2-(p-q)}r_{ij\cdot q+1,\cdots,p}/\sqrt{1-r^2_{ij\cdot q+1,\cdots,p}}$ 服从自由度为 $N-2-(p-q)$ 的 t 分布.

4.44　(4.3 节)　设 $\boldsymbol{X}'=(X_1, X_2, \boldsymbol{X}^{(2)\prime})$ 服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 给定 $X_2=x_2$ 和 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 X_1 的条件分布是

$$N\left[\mu_1+\gamma_2(x_2-\mu_2)+\boldsymbol{\gamma}'(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}), \sigma_{11\cdot 2,\cdots,p}\right],$$

其中

$$\begin{pmatrix} \sigma_{22} & \boldsymbol{\sigma}'_{(2)} \\ \boldsymbol{\sigma}_{(2)} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \gamma_2 \\ \boldsymbol{\gamma} \end{pmatrix}=\begin{pmatrix} \sigma_{12} \\ \boldsymbol{\sigma}_{(1)} \end{pmatrix}.$$

γ_2 和 $\boldsymbol{\gamma}$ 的估计定义为

$$\begin{pmatrix} a_{22} & \boldsymbol{a}'_{(2)} \\ \boldsymbol{a}_{(2)} & \boldsymbol{A}_{22} \end{pmatrix}\begin{pmatrix} c_2 \\ \boldsymbol{c} \end{pmatrix}=\begin{pmatrix} a_{12} \\ \boldsymbol{a}_{(1)} \end{pmatrix}.$$

证明 $c_2=a_{12\cdot 3,\cdots,p}/a_{22\cdot 3,\cdots,p}$. [提示: 依据 c_2 和那些 a 解出 $\boldsymbol{c}$, 再代入.]

4.45　(4.3 节)　用习题 4.44 的符号, 证明

$$\begin{aligned} a_{11\cdot 2,\cdots,p} &= a_{11}-\boldsymbol{a}'_{(1)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(1)}-c_2^2(a_{22}-\boldsymbol{a}'_{(2)}\boldsymbol{A}_{22}^{-1}\boldsymbol{a}_{(2)}) \\ &= a_{11\cdot 3,\cdots,p}-c_2^2 a_{22\cdot 3,\cdots,p}. \end{aligned}$$

提示: 利用

$$a_{11\cdot 2,\cdots,p}=a_{11}-\begin{pmatrix} c_2 & \boldsymbol{c}' \end{pmatrix}\begin{pmatrix} a_{22} & \boldsymbol{a}'_{(2)} \\ \boldsymbol{a}_{(2)} & \boldsymbol{A}_{22} \end{pmatrix}\begin{pmatrix} c_2 \\ \boldsymbol{c} \end{pmatrix}.$$

4.46　(4.3 节)　证明 $1/a_{22\cdot 3,\cdots,p}$ 是

$$\begin{pmatrix} a_{22} & \boldsymbol{a}'_{(2)} \\ \boldsymbol{a}_{(2)} & \boldsymbol{A}_{22} \end{pmatrix}^{-1}$$

左上角的元素.

4.47　(4.3 节)　利用习题 4.43~4.46 的结果, 证明 $\rho_{12\cdot 3,\cdots,p}=0$ 的检验等价于对 $\gamma_2=0$ 的普通 t 检验.

4.48 **缺失观测**. 设 $\boldsymbol{X} = (\boldsymbol{Y}' \quad \boldsymbol{Z}')'$ 服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 其中 $\boldsymbol{Y}$ 有 p 个分量, $\boldsymbol{Z}$ 有 q 个分量, 且有

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zz} \end{pmatrix}.$$

设 $\boldsymbol{X}$ 有 M 个观测, $\boldsymbol{Y}$ 上有另外 $N-M$ 个观测. 求 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的极大似然估计. [Anderson (1957).] [**提示**: 用 $\boldsymbol{Y}$ 的边缘密度和给定 $\boldsymbol{Y}$ 时 $\boldsymbol{Z}$ 的条件密度, 表达似然函数.]

4.49 假设 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, 其中

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}.$$

证明在一个观测 $\boldsymbol{x}' = (x_1, x_2, x_3)$ 的基础上, 利用

$$[x_2^2 + \chi_3^2(\alpha)]t^2 - 2(x_1x_2 + x_2x_3)t + x_1^2 + x_2^2 + x_3^2 - \chi_3^2(\alpha) = 0$$

关于 t 的解作为区间的端点, 我们可以得到 ρ 的一个置信区间 (置信度为 $1-\alpha$), 其中 $\chi_3^2(\alpha)$ 是自由度为 3 的 χ^2 分布在显著性水平为 α 的分位数.

第5章　广义 T^2 统计量

5.1　引　　言

在单变量统计学中有一类重要的问题, 即对于某个给定的分布, 当分布的方差未知时, 该分布的均值情况. 在样本的基础上, 人们希望判断其均值是否等于一个事先给定的值, 或希望给出这个均值存在的区间. 单变量统计学中常用的统计量, 是样本均值 $\bar{x}$ 与假设的总体均值 μ 之间的偏差除以样本标准差 s. 如果是从分布 $N(\mu,\sigma^2)$ 来抽样, 则

$$t=\sqrt{N}\frac{\bar{x}-\mu}{s} \tag{1}$$

服从众所周知的自由度为 $N-1$ 的 t 分布, 其中 N 是样本中观测的个数. 在这个事实上, 人们希望建立假设检验 $\mu=\mu_0$, 其中 μ_0 是给定的, 或者希望建立未知参数 μ 的置信区间.

在多元统计学中, 类比于 (1) 给出的 t 的平方是

$$T^2=N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}), \tag{2}$$

其中 $\bar{\boldsymbol{x}}$ 是容量为 N 的样本均值向量, $\boldsymbol{S}$ 是样本协方差阵. 我们将会说明这个统计量怎样用在检验关于总体均值向量 $\boldsymbol{\mu}$ 的假设, 或求这个未知 $\boldsymbol{\mu}$ 的置信区域. 在 (2) 式中的 $\boldsymbol{\mu}$ 等于被抽样的分布均值时和 $\boldsymbol{\mu}$ 不同于总体均值时, 分别求得 T^2 的分布. Hotelling (1931) 给出了两样本的 T^2 统计量及其当 $\boldsymbol{\mu}$ 是总体均值时的分布.

5.3 节展示了 T^2 的各种用法, 包括所有均值向量线性组合的联合置信区间. 给出了当 $\boldsymbol{\Sigma}$ 未知时的一个 James-Stein 估计. 5.4 节讨论了 T^2 检验的功效函数, 5.5 节讨论了多元 Behrens-Fisher 问题. 5.6 节考虑了 T^2 检验关于不变性和容许性的最优性质. 证明和应用了广义指数族的容许性 Stein 准则. 最后一节讨论了椭球等高分布的均值推断.

5.2　广义 T^2 统计量的推导及分布

5.2.1　作为似然比准则函数的 T^2 统计量的推导

T^2 统计量有很多用途, 我们首先说明, 基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本的假设 $H:\boldsymbol{\mu}=\boldsymbol{\mu}_0$, 其似然比检验就是基于 5.1 节的 (2) 式给出的 T^2 统计量. 假设我们有 N 个观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N(N>p)$. 似然比函数是

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma})=(2\pi)^{-\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})\right].\tag{1}$$

观测值是给定的. L 是不定元 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的一个函数. (我们不从符号中区分不定元和参数.) 似然比准则是

$$\lambda=\frac{\max_{\boldsymbol{\Sigma}}L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma})}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}}L(\boldsymbol{\mu},\boldsymbol{\Sigma})},\tag{2}$$

即分子是 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的似然函数在原假设限制的参数空间 ($\boldsymbol{\mu}=\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}$ 正定) 上的极大值, 分母是似然函数在整个参数空间 ($\boldsymbol{\Sigma}$ 正定的) 上的极大值. 当参数没有限制的时候, 极大值出现在当 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 为 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的极大似然估计时 (3.2 节),

$$\hat{\boldsymbol{\mu}}_\Omega=\bar{\boldsymbol{x}},\tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_\Omega=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'.\tag{4}$$

当 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ 时, 由引理 3.2.2 知, 似然函数在

$$\hat{\boldsymbol{\Sigma}}_\omega=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\mu}_0)(\boldsymbol{x}_\alpha-\boldsymbol{\mu}_0)'\tag{5}$$

有极大值. 而且由引理 3.2.2 知

$$\max_{\boldsymbol{\Sigma},\boldsymbol{\mu}}L(\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}}\mathrm{e}^{-\frac{1}{2}pN},\tag{6}$$

$$\max_{\boldsymbol{\Sigma}}L(\boldsymbol{\mu}_0,\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}\mathrm{e}^{-\frac{1}{2}pN}\tag{7}$$

因此似然比准则是

$$\begin{aligned}\lambda&=\frac{|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}}{|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}=\frac{|\sum(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'|^{\frac{1}{2}N}}{|\sum(\boldsymbol{x}_\alpha-\boldsymbol{\mu}_0)(\boldsymbol{x}_\alpha-\boldsymbol{\mu}_0)'|^{\frac{1}{2}N}}\\&=\frac{|\boldsymbol{A}|^{\frac{1}{2}N}}{|\boldsymbol{A}+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'|^{\frac{1}{2}N}},\end{aligned}\tag{8}$$

其中

$$\boldsymbol{A}=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'=(N-1)\boldsymbol{S}.\tag{9}$$

应用附录里的推论 A.3.1, 得到

$$\begin{aligned}\lambda^{2/N}&=\frac{|\boldsymbol{A}|}{\left|\boldsymbol{A}+\left[\sqrt{N}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\right]\left[\sqrt{N}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\right]'\right|}\\&=\frac{1}{1+N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{A}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)}\\&=\frac{1}{1+T^2/(N-1)},\end{aligned}\tag{10}$$

其中

$$T^2 = N(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0) = (N-1)N(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{A}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0). \tag{11}$$

似然比检验定义为下面的临界区域 (拒绝区域)

$$\lambda \leqslant \lambda_0, \tag{12}$$

其中 λ_0 的选取使得当原假设为真时 (12) 的概率等于显著性水平. 如果我们对 (12) 的两边同时取 $\frac{1}{2}N$ 次根, 再相继求倒数, 减去 1, 乘以 $N-1$, 得到

$$T^2 \geqslant T_0^2, \tag{13}$$

其中

$$T_0^2 = (N-1)(\lambda_0^{-2/N} - 1). \tag{14}$$

定理 5.2.1 对于 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, (13) 给出了假设检验 $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ 的似然比, 其中 T^2 定义为 (11) 式, $\bar{\boldsymbol{x}}$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的容量为 N 的样本均值, $\boldsymbol{S}$ 是样本的协方差阵, T_0^2 的选取使得当原假设为真时 (13) 的概率等于给定的显著性水平.

学生 t 检验有当 $\mu = 0$ 时关于比例变换不变的性质. 如果标量随机变量 X 服从分布 $N(\mu, \sigma^2)$, 则 $X^* = cX$ 服从分布 $N(c\mu, c^2\sigma^2)$, 它们属于同一类分布, 且 $\mathrm{E}(X) = 0$ 等价于 $\mathrm{E}(X^*) = \mathrm{E}(cX) = 0$. 如果观测 x_α 有相同形式的变换 ($x_\alpha^* = cx_\alpha$), 则对于 $c > 0$, 由 x_α^* 计算而得的 t^* 与由 x_α 计算而得的 t 相同. 因此, 无论测度的单位是什么, 其统计结果都是一样的.

广义 T^2 检验有相似的性质. 如果向量随机变量 $\boldsymbol{X}$ 服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 则 $\boldsymbol{X}^* = \boldsymbol{CX}$ (对 $|\boldsymbol{C}| \neq 0$) 服从分布 $N(\boldsymbol{C\mu}, \boldsymbol{C\Sigma C}')$, 它们也属于同一类分布. 假设 $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{0}$ 等价于假设 $\mathrm{E}(\boldsymbol{X}^*) = \mathrm{E}(\boldsymbol{CX}) = \boldsymbol{0}$. 如果观测 $\boldsymbol{x}_\alpha$ 有相同的变化 $\boldsymbol{x}_\alpha^* = \boldsymbol{C}\boldsymbol{x}_\alpha$, 则基于 $\boldsymbol{x}_\alpha^*$ 计算的 T^{*2} 等于基于 $\boldsymbol{x}_\alpha$ 计算的 T^2. 这是由 $\bar{\boldsymbol{x}}^* = \boldsymbol{C}\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A} = \boldsymbol{CAC}'$ 以及下面的引理得到的.

引理 5.2.1 对于任何 $p \times p$ 非奇异矩阵 $\boldsymbol{C}$ 和 $\boldsymbol{H}$ 以及任何向量 $\boldsymbol{k}$, 有

$$\boldsymbol{k}'\boldsymbol{H}^{-1}\boldsymbol{k} = (\boldsymbol{Ck})'(\boldsymbol{CHC}')^{-1}(\boldsymbol{Ck}). \tag{15}$$

证明 (15) 的右端项是

$$\begin{aligned}(\boldsymbol{Ck})'(\boldsymbol{CHC}')^{-1}(\boldsymbol{Ck}) &= \boldsymbol{k}'\boldsymbol{C}'(\boldsymbol{C}')^{-1}\boldsymbol{H}^{-1}\boldsymbol{C}^{-1}\boldsymbol{Ck} \\ &= \boldsymbol{k}'\boldsymbol{H}^{-1}\boldsymbol{k}.\end{aligned} \tag{16}$$

■

我们将在 5.6 节证明, 在所有关于此类变换不变的检验中, (13) 是一致最优的. 我们可以从超平行体的角度, 给出似然准则的 $\frac{1}{2}N$ 次根

$$\lambda^{2/N} = \frac{|\sum_{\alpha=1}^N (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'|}{|\sum_{\alpha=1}^N (\boldsymbol{x}_\alpha - \boldsymbol{\mu}_0)(\boldsymbol{x}_\alpha - \boldsymbol{\mu}_0)'|} \tag{17}$$

的几何解释. (见 7.5 节.) 在 p 维表示中 $\lambda^{2/N}$ 的分子是所有主边为 p 维向量, 一个

端点在 $\bar{\boldsymbol{x}}$, 另一个端点在 $\boldsymbol{x}_\alpha$ 的超平行体的体积平方和. 分母是所有主边为 p 维向量, 一个端点在 $\boldsymbol{\mu}_0$, 另一个端点在 $\boldsymbol{x}_\alpha$ 的超平行体的体积平方和. 如果包含以 $\bar{\boldsymbol{x}}$($\boldsymbol{x}_\alpha$ 的中心) 为"中心"发散的向量的体积平方和, 远小于包含以 $\boldsymbol{\mu}_0$ 为中心发散的向量的体积平方和, 则我们拒绝 $\boldsymbol{\mu}_0$ 是分布的均值的假设.

在 N 维空间里也有一个解释. 设 $\boldsymbol{y}_i=(x_{i1},\cdots,x_{iN})'$ 是第 i 个向量. 则

$$\sqrt{N}\bar{x}_i=\sum_{\alpha=1}^{N}\frac{1}{\sqrt{N}}x_{i\alpha} \tag{18}$$

是 $\boldsymbol{y}_i$ 在等角线 (有方向余弦 $1/\sqrt{N},\cdots,1/\sqrt{N}$) 上的投影到原点的距离. 投影的坐标是 $(\bar{x}_i,\cdots,\bar{x}_i)$. 则 $(x_{i1}-\bar{x}_i,\cdots,x_{iN}-\bar{x}_i)$ 是 y_i 在过原点且垂直于等角线的平面上的投影. $\lambda^{2/N}$ 的分子是主边为向量 $(x_{i1}-\bar{x}_i,\cdots,x_{iN}-\bar{x}_i)$ 的超平行体的 p 维体积的平方. 点 $(x_{i1}-\mu_{0i},\cdots,x_{iN}-\mu_{0i})$ 是由 $\boldsymbol{y}_i$ 平移 (距离为 $\sqrt{N}\mu_{0i}$) 到等角线上得到的. $\lambda^{2/N}$ 的分母是以这些向量为主边的超平行体的体积平方. 则 $\lambda^{2/N}$ 是这两个平方体积的比.

5.2.2 T^2 的分布

本小节我们将在一般条件下求 T^2 的分布, 包括当原假设不真的情况. 设 $T^2=\boldsymbol{Y}'\boldsymbol{S}^{-1}\boldsymbol{Y}$, 其中 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{\nu},\boldsymbol{\Sigma})$, $n\boldsymbol{S}$ 与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 独立且同分布, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 独立服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. 5.2.1 节中定义的 T^2 是取 $\boldsymbol{Y}=\sqrt{N}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$ 和 $\boldsymbol{\nu}=\sqrt{N}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$ 以及 $n=N-1$ 时的特殊情况. 设 $\boldsymbol{D}$ 是一个非奇异矩阵, 满足 $\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}'=\boldsymbol{I}$, 定义

$$\boldsymbol{Y}^*=\boldsymbol{D}\boldsymbol{Y},\quad \boldsymbol{S}^*=\boldsymbol{D}\boldsymbol{S}\boldsymbol{D}',\quad \boldsymbol{\nu}^*=\boldsymbol{D}\boldsymbol{\nu}. \tag{19}$$

则 $T^2=\boldsymbol{Y}^{*\prime}\boldsymbol{S}^{*-1}\boldsymbol{Y}^*$(引理 5.2.1), 其中 $\boldsymbol{Y}^*$ 服从分布 $N(\boldsymbol{\nu}^*,\boldsymbol{I})$, $n\boldsymbol{S}^*$ 与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha^*\boldsymbol{Z}_\alpha^{*\prime}=\sum_{\alpha=1}^{n}\boldsymbol{D}\boldsymbol{Z}_\alpha\boldsymbol{D}\boldsymbol{Z}_\alpha{}'$ 独立且同分布, 其中 $\boldsymbol{Z}_\alpha^*=\boldsymbol{D}\boldsymbol{Z}_\alpha$ 独立同服从分布 $N(\boldsymbol{0},\boldsymbol{I})$. 由引理 5.2.1, 我们有 $\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}=\boldsymbol{\nu}^{*\prime}(\boldsymbol{I})^{-1}\boldsymbol{\nu}^*=\boldsymbol{\nu}^{*\prime}\boldsymbol{\nu}^*$.

定义 $p\times p$ 正交矩阵 $\boldsymbol{Q}$ 的第一行为

$$q_{1i}=\frac{Y_i^*}{\sqrt{\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*}},\quad i=1,\cdots,p, \tag{20}$$

这是合理的, 因为 $\sum_{i=1}^{p}q_{1i}^2=1$. 其他 $p-1$ 行可以由某随意的法则 (附录中引理 A.4.2) 定义. 因为 $\boldsymbol{Q}$ 依赖 $\boldsymbol{Y}^*$, 所以它是一个随机矩阵. 现在设

$$\boldsymbol{U}=\boldsymbol{Q}\boldsymbol{Y}^*,\quad \boldsymbol{B}=\boldsymbol{Q}n\boldsymbol{S}^*\boldsymbol{Q}'. \tag{21}$$

由 $\boldsymbol{Q}$ 的定义,

$$U_1=\sum q_{1i}Y_i^*=\sqrt{\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*},\quad U_j=\sum q_{ji}Y_i^*=\sqrt{\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*}\sum q_{ji}q_{1i}=0,\quad j\neq 1. \tag{22}$$

则

$$\frac{T^2}{n} = \boldsymbol{U}'\boldsymbol{B}^{-1}\boldsymbol{U} = (U_1, 0, \cdots, 0)\begin{pmatrix} b^{11} & b^{12} & \cdots & b^{1p} \\ b^{21} & b^{22} & \cdots & b^{2p} \\ \vdots & \vdots & & \vdots \\ b^{p1} & b^{p2} & \cdots & b^{pp} \end{pmatrix}\begin{pmatrix} U_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{23}$$

$$= U_1^2 b^{11},$$

其中 $(b^{ij}) = \boldsymbol{B}^{-1}$. 由附录中定理 A.3.3 知, $1/b^{11} = b_{11} - \boldsymbol{b}'_{(1)}\boldsymbol{B}_{22}^{-1}\boldsymbol{b}_{(1)} = b_{11\cdot 2,\cdots,p}$, 其中

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & \boldsymbol{b}'_{(1)} \\ \boldsymbol{b}_{(1)} & \boldsymbol{B}_{22} \end{pmatrix}, \tag{24}$$

$T^2/n = U_1^2/b_{11\cdot 2,\cdots,p} = \boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*/b_{11\cdot 2,\cdots,p}$. 给定 $\boldsymbol{Q}$ 时 $\boldsymbol{B}$ 的条件分布与 $\sum_{\alpha=1}^{n}\boldsymbol{V}_\alpha\boldsymbol{V}'_\alpha$ 的分布相同, 其中需附加条件 $\boldsymbol{V}_\alpha = \boldsymbol{Q}\boldsymbol{Z}^*_\alpha$ 独立服从分布 $N(\boldsymbol{0}, \boldsymbol{I})$. 由定理 4.3.3, $b_{11\cdot 2,\cdots,p}$ 是条件地与 $\sum_{\alpha=1}^{n-(p-1)} W_\alpha^2$ 同分布, 其中需附加条件 W_α 独立服从分布 $N(0,1)$, 即 $b_{11\cdot 2,\cdots,p}$ 条件地服从自由度为 $n-(p-1)$ 的 χ^2 分布. 因为 $b_{11\cdot 2,\cdots,p}$ 的条件分布不依赖 $\boldsymbol{Q}$, 所以它是无条件地服从 χ^2 分布. $\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*$ 服从非中心 χ^2 分布, 其自由度为 p, 非中心参数为 $\boldsymbol{\nu}^{*\prime}\boldsymbol{\nu}^* = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$. 则 T^2/n 的分布如同一个非中心 χ^2 分布与另一个独立 χ^2 分布的比.

定理 5.2.2　*设 $T^2 = \boldsymbol{Y}'\boldsymbol{S}^{-1}\boldsymbol{Y}$, 其中 $\boldsymbol{Y}$ 服从分布 $N(\boldsymbol{\nu}, \boldsymbol{\Sigma})$, $n\boldsymbol{S}$ 与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 独立且同分布, 其中 $\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n$ 独立服从分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. 则 $(T^2/n)[(n-p+1)/p]$ 服从自由度为 p 和 $n-p+1$, 非中心参数为 $\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$ 的非中心 F 分布. 如果 $\boldsymbol{\nu} = \boldsymbol{0}$, 则是中心 F.*

我们称此为自由度为 n 的 T^2 分布.

推论 5.2.1　*设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的样本, 并设 $T^2 = N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$. $[T^2/(N-1)][(N-p)/p]$ 的分布是自由度为 p 和 $N-p$, 非中心参数为 $N(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)$ 的非中心 F 分布. 如果 $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, 则此 F 分布是中心化的.*

上面关于 T^2 分布的导出是 Bowker (1960) 给出的. 5.4 节讨论了非中心 F 密度及其分布表.

对于大样本, 即使母分布不是正态的, 推论 5.2.1 定义的 T^2 分布也是近似有效的, 从这种角度讲, T^2 检验是一个稳健方法.

定理 5.2.3　*设 $\{\boldsymbol{X}_\alpha\}, \alpha = 1, 2, \cdots,$ 是一列独立同分布的随机向量, 其均值向量为 $\boldsymbol{\mu}$, 协方差阵为 $\boldsymbol{\Sigma}$; 设 $\bar{\boldsymbol{X}}_N = (1/N)\sum_{\alpha=1}^{N}\boldsymbol{X}_\alpha, \boldsymbol{S}_N = [1/(N-1)]\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}}_N)(\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}}_N)'$ 以及 $T_N^2 = N(\bar{\boldsymbol{X}}_N - \boldsymbol{\mu}_0)'\boldsymbol{S}_N^{-1}(\bar{\boldsymbol{X}}_N - \boldsymbol{\mu}_0)$. 则当 $N \to \infty$ 时, 如果 $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, 那么 T_N^2 的极限分布是自由度为 p 的 χ^2 分布.*

证明　由中心极限定理 (定理 4.2.3), $\sqrt{N}(\bar{\boldsymbol{X}}_N - \boldsymbol{\mu})$ 的极限分布是 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. 样本协方差阵随机收敛到 $\boldsymbol{\Sigma}$. 则当 $N \to \infty$ 时 T_N^2 的极限分布是 $\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$ 的分布,

其中 $\boldsymbol{Y}$ 服从分布 $N(\mathbf{0},\boldsymbol{\Sigma})$. 由定理 3.3.3 得到此定理. ■

当原假设为真时, T^2/n 的分布如同 χ_p^2/χ_{n-p+1}^2 的分布, 且由 (10) 给出的 $\lambda^{2/N}$ 的分布如同 $\chi_{n-p+1}^2/(\chi_{n-p+1}^2+\chi_p^2)$ 的分布. 当 χ_a^2 和 χ_b^2 独立时, $V=\chi_a^2/(\chi_a^2+\chi_b^2)$ 的密度是

$$\frac{\Gamma[\frac{1}{2}(a+b)]}{\Gamma(\frac{1}{2}a)\Gamma(\frac{1}{2}b)}v^{\frac{1}{2}a-1}(1-v)^{\frac{1}{2}b-1}=\beta\left(v;\frac{1}{2}a,\frac{1}{2}b\right); \tag{25}$$

这是参数为 $\frac{1}{2}a$ 和 $\frac{1}{2}b$ 的β 分布的密度 (习题 5.27). 因此, $\lambda^{2/N}=(1+T^2/n)^{-1}$ 的分布是参数为 $\frac{1}{2}p$ 和 $\frac{1}{2}(n-p+1)$ 的 β 分布.

5.3 T^2 统计量的应用

5.3.1 关于均值向量为给定向量的假设检验

基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本, 假设 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ 的似然比检验等价于 5.2.1 节给出的

$$T^2\geqslant T_0^2. \tag{1}$$

如果显著性水平是 α, 则得到 F 分布的 $100\alpha\%$ 的点, 即

$$T_0^2=\frac{(N-1)p}{N-p}F_{p,N-p}(\alpha)=T_{p,N-1}^2(\alpha). \tag{2}$$

显著性水平的选取可能依赖检验的功效. 我们将在 5.4 节讨论.

统计量 T^2 由 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}$ 计算而得. 向量 $\boldsymbol{A}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)=\boldsymbol{b}$ 是 $\boldsymbol{A}\boldsymbol{b}=\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0$ 的解. 则 $T^2/(N-1)=N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{b}$.

注意 $T^2/(N-1)$ 是

$$|N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'-\lambda\boldsymbol{A}|=0 \tag{3}$$

的非零根.

引理 5.3.1 如果 $\boldsymbol{\nu}$ 是一个 p 维向量并且 $\boldsymbol{B}$ 是一个非奇异 $p\times p$ 矩阵, 则 $\boldsymbol{\nu}'\boldsymbol{B}^{-1}\boldsymbol{\nu}$ 是

$$|\boldsymbol{\nu}\boldsymbol{\nu}'-\lambda\boldsymbol{B}|=0 \tag{4}$$

的非零根.

证明 (4) 的非零根, 比如说 λ_1, 与一个特征向量 $\boldsymbol{\beta}$ 有关, 满足

$$\boldsymbol{\nu}\boldsymbol{\nu}'\boldsymbol{\beta}=\lambda_1\boldsymbol{B}\boldsymbol{\beta}. \tag{5}$$

因为 $\lambda_1\neq 0$, 则 $\boldsymbol{\nu}'\boldsymbol{\beta}\neq\mathbf{0}$. 左乘 $\boldsymbol{\nu}'\boldsymbol{B}^{-1}$, 我们得到

$$(\boldsymbol{\nu}'\boldsymbol{B}^{-1}\boldsymbol{\nu})(\boldsymbol{\nu}'\boldsymbol{\beta})=\lambda_1(\boldsymbol{\nu}'\boldsymbol{\beta}). \quad ■ \tag{6}$$

在上面的情况里, $\boldsymbol{\nu}=\sqrt{N}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$ 和 $\boldsymbol{B}=\boldsymbol{A}$.

5.3.2 均值向量的置信区域

如果 $\boldsymbol{\mu}$ 是 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的均值, 若要抽取容量为 N 的样本, 其均值 $\bar{x}$ 和协方差阵

S 满足

$$N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \leqslant T^2_{p,N-1}(\alpha), \tag{7}$$

则取到该样本的概率是 $1-\alpha$. 因此, 如果我们对特定样本计算 (7), 则 (7) 关于 $\boldsymbol{\mu}$ 为真的置信度为 $1-\alpha$. 不等式

$$N(\bar{\boldsymbol{x}}-\boldsymbol{m})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{m}) \leqslant T^2_{p,N-1}(\alpha) \tag{8}$$

是 $\boldsymbol{m}$ 的 p 维空间中一个椭球的内部和边界, 这个椭球的中心在 $\bar{\boldsymbol{x}}$ 上, 大小和形状依赖 $\boldsymbol{S}^{-1}$ 和 α. 见图 5.1. 我们说 $\boldsymbol{\mu}$ 以置信度 $1-\alpha$ 落在这个椭球. 对于随机抽样, (8) 是一个随机椭球.

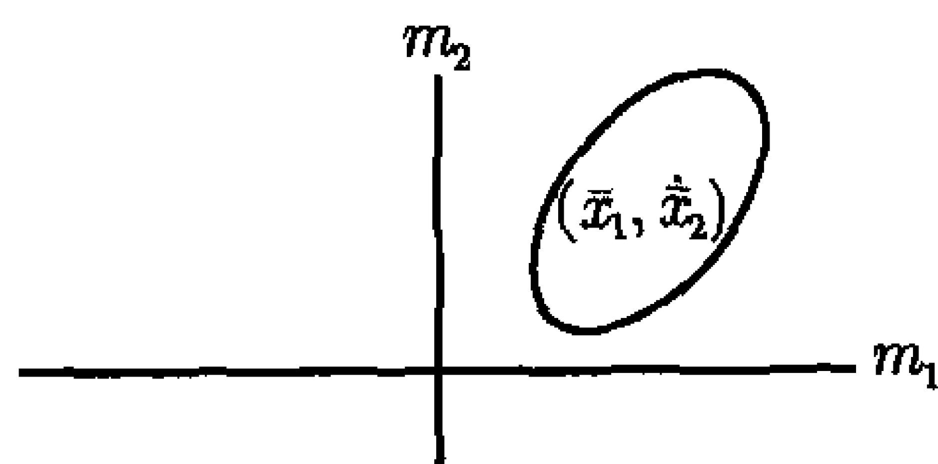


图 5.1 一个置信椭圆

5.3.3 所有均值向量线性组合的联合置信区间

从 $\boldsymbol{\mu}$ 的置信区域 (8), 我们可以得到线性函数 $\boldsymbol{\gamma}'\boldsymbol{\mu}$ 在给定置信度下同时存在的置信区间.

引理 5.3.2 (广义柯西–施瓦茨不等式) 对于正定矩阵 $\boldsymbol{S}$, 有

$$(\boldsymbol{\gamma}'\boldsymbol{y})^2 \leqslant \boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}\boldsymbol{y}'\boldsymbol{S}^{-1}\boldsymbol{y}. \tag{9}$$

证明 设 $b=\boldsymbol{\gamma}'\boldsymbol{y}/\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}$. 则

$$\begin{aligned}0 &\leqslant (\boldsymbol{y}-b\boldsymbol{S}\boldsymbol{\gamma})'\boldsymbol{S}^{-1}(\boldsymbol{y}-b\boldsymbol{S}\boldsymbol{\gamma}) \\ &= \boldsymbol{y}'\boldsymbol{S}^{-1}\boldsymbol{y} - b\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{\gamma}b + b^2\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{S}\boldsymbol{\gamma} \\ &= \boldsymbol{y}'\boldsymbol{S}^{-1}\boldsymbol{y} - \frac{(\boldsymbol{\gamma}'\boldsymbol{y})^2}{\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}},\end{aligned} \tag{10}$$

即得到 (9). ■

当 $\boldsymbol{y}=\bar{\boldsymbol{x}}-\boldsymbol{\mu}$ 时, 则 (9) 意味着

$$\begin{aligned}|\boldsymbol{\gamma}'(\bar{\boldsymbol{x}}-\boldsymbol{\mu})| &\leqslant \sqrt{\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})} \\ &\leqslant \sqrt{\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}}\sqrt{T^2_{p,N-1}(\alpha)/N}\end{aligned} \tag{11}$$

以概率 $1-\alpha$ 对所有的 $\boldsymbol{\gamma}$ 成立. 因此, 我们可以断定未知参数向量以置信度 $1-\alpha$, 对所有的 $\boldsymbol{\gamma}$ 同时满足不等式

$$|\boldsymbol{\gamma}'\bar{\boldsymbol{x}}-\boldsymbol{\gamma}'\boldsymbol{m}| \leqslant \sqrt{\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}}\sqrt{T^2_{p,N-1}(\alpha)/N}. \tag{12}$$

置信区域 (8) 可以通过设置 (12) 中的 $\boldsymbol{\gamma}$ 为如 $(1,0,\cdots,0)'$ 的简单向量得到 m_1, 如 $(1,-1,0,\cdots,0)$ 得到 m_1-m_2, 等等. 需要注意, 如果只对一个线性函数 $\boldsymbol{\gamma}'\boldsymbol{\mu}$ 感兴趣, 则可用 $t_n(\alpha)$ 代替 $\sqrt{T^2_{p,N-1}(\alpha)}=\sqrt{npF_{p,n-p+1}(\alpha)/(n-p+1)}$.

5.3.4 两样本问题

T^2 统计量的另一个应用是当两个正态总体的协方差阵未知且相等时, 关于它们均值相等的原假设. 假设 $\boldsymbol{y}_1^{(i)},\cdots,\boldsymbol{y}_{N_i}^{(i)}$ 是来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})$ $(i=1,2)$ 的样本. 我们

想检验原假设 $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$. 向量 $\bar{\boldsymbol{y}}^{(i)}$ 服从分布 $N[\boldsymbol{\mu}^{(i)}, (1/N_i)\boldsymbol{\Sigma}]$. 从而在原假设下, $\sqrt{N_1N_2/(N_1+N_2)}(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)})$ 服从分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. 如果我们设

$$\begin{aligned}\boldsymbol{S} = \frac{1}{N_1+N_2-2}&\left\{\sum_{\alpha=1}^{N_1}(\boldsymbol{y}_\alpha^{(1)} - \bar{\boldsymbol{y}}^{(1)})(\boldsymbol{y}_\alpha^{(1)} - \bar{\boldsymbol{y}}^{(1)})'\right. \\ &\left.+\sum_{\alpha=1}^{N_2}(\boldsymbol{y}_\alpha^{(2)} - \bar{\boldsymbol{y}}^{(2)})(\boldsymbol{y}_\alpha^{(2)} - \bar{\boldsymbol{y}}^{(2)})'\right\},\end{aligned} \tag{13}$$

则 $(N_1+N_2-2)\boldsymbol{S}$ 和 $\sum_{\alpha=1}^{N_1+N_2-2}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 同分布, 其中 $\boldsymbol{Z}_\alpha$ 服从分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. 因此

$$T^2 = \frac{N_1N_2}{N_1+N_2}(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)}) \tag{14}$$

和自由度为 N_1+N_2-2 的 T^2 分布相同. 显著性水平为 α 的临界区域是

$$T^2 > \frac{(N_1+N_2-2)p}{N_1+N_2-p-1}F_{p,N_1+N_2-p-1}(\alpha). \tag{15}$$

$\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ 的一个置信度为 $1-\alpha$ 的置信区域是满足

$$\begin{aligned}&(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)} - \boldsymbol{m})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)} - \boldsymbol{m}) \\ \leqslant &\frac{N_1+N_2}{N_1N_2}T^2_{p,N_1+N_2-2}(\alpha) \\ = &\frac{N_1+N_2}{N_1N_2}\frac{(N_1+N_2-2)p}{N_1+N_2-p-1}F_{p,N_1+N_2-p-1}(\alpha)\end{aligned} \tag{16}$$

的向量 $\boldsymbol{m}$ 的集合. 联合置信区间是

$$|\boldsymbol{\gamma}'(\bar{\boldsymbol{y}}^{(1)} - \bar{\boldsymbol{y}}^{(2)}) - \boldsymbol{\gamma}'\boldsymbol{m}| \leqslant \sqrt{\boldsymbol{\gamma}'\boldsymbol{S}\boldsymbol{\gamma}}\sqrt{\frac{N_1+N_2}{N_1N_2}T^2_{p,N_1+N_2-2}(\alpha)}. \tag{17}$$

考虑 Fisher (1936) 中的一个例子. 设 x_1 = 萼片长度, x_2 = 萼片宽度, x_3 = 花瓣长度, x_4 = 花瓣宽度. 总体变色鸢尾 (1) 有 50 个观测, 总体山鸢尾 (2) 有 50 个观测. 见表 3.4. 整理这些数据 (单位厘米) 得到

$$\bar{\boldsymbol{x}}^{(1)} = \begin{pmatrix} 5.936 \\ 2.770 \\ 4.260 \\ 1.326 \end{pmatrix}, \tag{18}$$

$$\bar{\boldsymbol{x}}^{(2)} = \begin{pmatrix} 5.006 \\ 3.428 \\ 1.462 \\ 0.246 \end{pmatrix}, \tag{19}$$

$$98\boldsymbol{S}=\begin{pmatrix} 19.1434 & 9.0356 & 9.7634 & 3.2394 \\ 9.0356 & 11.8658 & 4.6232 & 2.4746 \\ 9.7634 & 4.6232 & 12.2978 & 3.8794 \\ 3.2394 & 2.4746 & 3.8794 & 2.4604 \end{pmatrix}. \tag{20}$$

$T^2/98$ 的值是 26.334, 从而 $(T^2/98)\times\frac{95}{4}=625.5$. 这个值在显著性水平为 0.01 时相比于自由度为 4 和 95 的 F 值 3.52 来说, 是非常显著的.

不同均值分量差的联合置信区间 $\mu_i^{(1)}-\mu_i^{(2)}(i=1,2,3,4)$ 是 0.930 ± 0.337, -0.658 ± 0.265, -2.708 ± 0.270 和 1.080 ± 0.121. 在每一种情况, 0 都不在区间内. [因为 $t_{98}(0.01)<T_{4,98}(0.01)$, 所以任何分量上的单变量检验都会导致拒绝原假设.] 最后两分量说明其与 0 的差最显著.

5.3.5 多样本问题

思考过上面的例子后, Fisher 又考虑了一个来自有相同协方差阵的总体的第三个样本. 他在 50 个维吉尼亚鸢尾上做相同测量 (表 3.4). 理论认为这 3 个物种的基因结构使得其均值向量有关系

$$3\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(3)}+2\boldsymbol{\mu}^{(2)}, \tag{21}$$

其中, $\boldsymbol{\mu}^{(3)}$ 是第三个总体的均值向量.

这是下面一般问题的特殊情况. 设 $\{\boldsymbol{x}_\alpha^{(i)}\}(\alpha=1,\cdots,N_i, i=1,\cdots,q)$ 是分别来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})(i=1,\cdots,q)$ 的样本. 我们检验假设

$$H:\sum_{i=1}^q\beta_i\boldsymbol{\mu}^{(i)}=\boldsymbol{\mu}, \tag{22}$$

其中 $\beta_1,\cdots,\beta_q$ 是给定的标量, $\boldsymbol{\mu}$ 是给定的向量. 准则是

$$T^2=c\left(\sum_{i=1}^q\beta_i\bar{\boldsymbol{x}}^{(i)}-\boldsymbol{\mu}\right)'\boldsymbol{S}^{-1}\left(\sum_{i=1}^q\beta_i\bar{\boldsymbol{x}}^{(i)}-\boldsymbol{\mu}\right), \tag{23}$$

其中

$$\bar{\boldsymbol{x}}^{(i)}=\frac{1}{N_i}\sum_{\alpha=1}^{N_i}\boldsymbol{x}_\alpha^{(i)}, \tag{24}$$

$$\left(\sum_{i=1}^q N_i-q\right)\boldsymbol{S}=\sum_{i=1}^q\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})', \tag{25}$$

$$\frac{1}{c}=\sum_{i=1}^q\frac{\beta_i^2}{N_i}. \tag{26}$$

这个 T^2 服从自由度为 $\sum_{i=1}^q N_i-q$ 的 T^2 分布.

实际上, Fisher 在他的例子中假定这 3 个总体的协方差阵可以不同. 他用到了 5.5 节的方法.

5.3.6 对称性问题

考虑假设检验 $H:\mu_1=\mu_2=\cdots=\mu_p$, 它基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本 $x_1,\cdots,x_N$, 其中 $\boldsymbol{\mu}'=(\mu_1,\cdots,\mu_p)$. 设 $\boldsymbol{C}$ 是任意满足

$$\boldsymbol{C}\boldsymbol{\varepsilon}=\mathbf{0} \tag{27}$$

的秩为 $p-1$ 的 $(p-1)\times p$ 矩阵, 其中 $\varepsilon'=(1,\cdots,1)$. 则

$$\boldsymbol{y}_\alpha=\boldsymbol{C}\boldsymbol{x}_\alpha,\quad \alpha=1,\cdots,N \tag{28}$$

有均值 $\boldsymbol{C}\boldsymbol{\mu}$ 和协方差阵 $\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$. 假设 H 是 $\boldsymbol{C}\boldsymbol{\mu}=\mathbf{0}$. 使用的统计量是

$$T^2=N\bar{\boldsymbol{y}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{y}}, \tag{29}$$

其中

$$\bar{\boldsymbol{y}}=\frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{y}_\alpha=\boldsymbol{C}\bar{\boldsymbol{x}}, \tag{30}$$

$$\begin{aligned}\boldsymbol{S}&=\frac{1}{N-1}\sum_{\alpha=1}^{N}(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})' \\ &=\frac{1}{N-1}\boldsymbol{C}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'\boldsymbol{C}'.\end{aligned} \tag{31}$$

这个统计量服从自由度为 $N-1$ 的 $p-1$ 维 T^2 分布. 这个 T^2 统计量在与 ε 正交的 $p-1$ 维空间中的线性变换下是不变的. 因此这个统计量独立于 $\boldsymbol{C}$ 的选择.

Rao (1948b) 给出了一个此类例子. 设 N 是对一棵软木树从北面钻木得到的软木量, 类似地定义 E, S 和 W. 一棵树上的 4 个钻木量可以看做来自一个 4 元正态分布的观测. 问题是软木树在每个面的钻木量是否都相同. 我们做变换

$$\begin{aligned}y_1&=N-E-W+S,\\ y_2&=S-W,\\ y_3&=N-S.\end{aligned} \tag{32}$$

观测的个数是 28. 均值向量是

$$\bar{\boldsymbol{y}}=\begin{pmatrix}8.86\\4.50\\0.86\end{pmatrix}, \tag{33}$$

$\boldsymbol{y}$ 的协方差阵是

$$\boldsymbol{S}=\begin{pmatrix}128.72 & 61.41 & -21.02\\ 61.41 & 56.93 & -28.30\\ -21.02 & -28.30 & 63.53\end{pmatrix}. \tag{34}$$

$T^2/(N-1)$ 的值是 0.768. 则在 1% 水平下, 和自由度为 3 和 25 的 F 分位数相比, 统计量 $0.768\times 25/3=6.402$ 是显著的.

5.3.7　均值的改良估计

在 3.5 节中, 我们考虑了协方差阵已知时均值的估计, 并证明了基于此的施坦型估计比样本均值可产生更小的二次风险. 特别地, 如果损失为 $(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{\mu})$, 则对任意 $\boldsymbol{\nu}$,

$$\left(1-\frac{p-2}{N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})}\right)^{+}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})+\boldsymbol{\nu} \tag{35}$$

是 $\boldsymbol{\mu}$ 的一个极小化极大估计, 并且当 $p \geqslant 3$ 时有比 $\bar{\boldsymbol{x}}$ 更小的风险. 当 $\boldsymbol{\Sigma}$ 未知时, 我们考虑用它的一个估计代替, 也就是, $\boldsymbol{A}=n\boldsymbol{S}$ 的一个倍数.

定理 5.3.1　当损失函数是 $(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{\mu})$ 时, 对于 $p \geqslant 3$, 由

$$\left(1-\frac{a}{N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})}\right)(\bar{\boldsymbol{x}}-\boldsymbol{\nu})+\boldsymbol{\nu} \tag{36}$$

给出的估计比 $\bar{\boldsymbol{x}}$ 有更小的风险, 并且当 $0<a<2(p-2)/(n-p+3)$ 时是极小化极大的, 风险在 $a=(p-2)/(n-p+3)$ 达到极小值.

证明　如 $\boldsymbol{\Sigma}$ 已知时的情况 (3.5.2 节), 我们可以做变换使得 $(1/N)\boldsymbol{\Sigma}$ 变换为 $\boldsymbol{I}$. 则问题变为基于分布为 $N(\boldsymbol{\mu},\boldsymbol{I})$ 的 $\boldsymbol{Y}$ 去估计 $\boldsymbol{\mu}$ 且 $\boldsymbol{A}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 独立服从分布 $N(\boldsymbol{0},\boldsymbol{I})$, 损失函数是 $(\boldsymbol{m}-\boldsymbol{\mu})'(\boldsymbol{m}-\boldsymbol{\mu})$. (我们去掉一个 N 的因子.) 风险差是

$$\begin{aligned}\Delta R(\boldsymbol{\mu})&=\mathrm{E}_{\boldsymbol{\mu}}\left\{\|\boldsymbol{Y}-\boldsymbol{\mu}\|^2-\left\|\left(1-\frac{a}{(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{\nu})}\right)(\boldsymbol{Y}-\boldsymbol{\nu})+\boldsymbol{\nu}-\boldsymbol{\mu}\right\|^2\right\} \\ &=\mathrm{E}_{\boldsymbol{\mu}}\left\{2\frac{a}{(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{\nu})}\sum_{i=1}^{p}(Y_i-\mu_i)(Y_i-\nu_i)\right. \\ &\quad\left.-\frac{a^2}{[(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{\nu})]^2}\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2\right\}.\end{aligned} \tag{37}$$

定理 5.2.2 的证明说明了 $(\boldsymbol{Y}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\boldsymbol{Y}-\boldsymbol{\nu})$ 的分布如同 $\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2/\chi^2_{n-p+1}$ 的分布, 其中 χ^2_{n-p+1} 独立于 $\boldsymbol{Y}$. 则风险差是

$$\begin{aligned}\Delta R(\boldsymbol{\mu})&=\mathrm{E}_{\boldsymbol{\mu}}\left\{\frac{2a\chi^2_{n-p+1}}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\sum_{i=1}^{p}(Y_i-\mu_i)(Y_i-\nu_i)-\frac{a^2(\chi^2_{n-p+1})^2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\right\} \\ &=\mathrm{E}_{\boldsymbol{\mu}}\left\{\frac{2a(p-2)\chi^2_{n-p+1}}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}-\frac{a^2(\chi^2_{n-p+1})^2}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}\right\} \\ &=\{2(p-2)(n-p+1)a \\ &\quad-[2(n-p+1)+(n-p+1)^2]a^2\}\,\mathrm{E}_{\boldsymbol{\mu}}\frac{1}{\|\boldsymbol{Y}-\boldsymbol{\nu}\|^2}.\end{aligned} \tag{38}$$

花括号里的因子是 $n-p+1$ 乘以 $2(p-2)a-(n-p+3)a^2$, 当 $0<a<2(p-2)/(n-p+3)$ 时它是正的, 当 $a=(p-2)/(n-p+3)$ 时它取极大值. ■

在 $\boldsymbol{Y}$ 的风险上的改进是 $(n-p+1)(p-2)^2/(n-p+3)\cdot \mathrm{E}_{\boldsymbol{\mu}}\|\boldsymbol{Y}-\boldsymbol{\nu}\|^{-2}$, 而当 $\boldsymbol{\Sigma}$ 已知时, 3.5 节中 $\boldsymbol{m}(\boldsymbol{y})$ 的改进 $(p-2)^2\mathrm{E}_{\boldsymbol{\mu}}\|\boldsymbol{Y}-\boldsymbol{\nu}\|^{-2}$.

推论 5.3.1 当 $p\geqslant 3$ 时, 估计

$$\left(1-\frac{a}{N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{A}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})}\right)^{+}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})+\boldsymbol{\nu} \tag{39}$$

有比 (36) 更小的风险且在 $0<a<2(p-2)/(n-p+3)$ 时是极小化极大的.

证明 这个推论可由定理 5.3.1 和引理 3.5.2 得出. ∎

(39) 的风险不必在 $a=(p-2)/(n-p+3)$ 极小化, 但是这个值看上去是个不错的选择. 这是将 3.5 节的估计 (18) 中的 $\boldsymbol{\Sigma}$ 替换为 $[1/(n-p+3)]\boldsymbol{A}$.

当损失函数为 $(\boldsymbol{m}-\boldsymbol{\mu})'\boldsymbol{Q}(\boldsymbol{m}-\boldsymbol{\mu})$ 时, 其中 $\boldsymbol{Q}$ 是正定矩阵, 很难找到一个好的一致改良估计. 可以替换 3.5 节的估计量中的 $\boldsymbol{\Sigma}$ 为其估计.

5.4 备择假设下 T^2 的分布, 功效函数

在 5.2.2 节我们证明了 $(T^2/n)(N-p)/p$ 服从一个非中心 F 分布. 本节我们将讨论这个非中心 F 分布、它的表及其在基于 T^2 的方法上的应用.

这个非中心 F 分布定义为一个非中心 χ^2 和一个独立的 χ^2 除以相应的自由度的比的分布. 设 V 是服从自由度为 p, 非中心参数为 τ^2 的 χ^2 分布 (如定理 3.3.5 所给), 设 W 独立服从自由度为 m 的 χ^2 分布. 我们可求 $F=(V/p)/(W/m)$ 的密度, 它是非中心参数为 τ^2 的非中心 F. V 和 W 的联合密度是 3.3 节的 (28) 乘以 W 的密度 $2^{-\frac{1}{2}m}\Gamma^{-1}\left(\frac{1}{2}m\right)w^{\frac{1}{2}m-1}\mathrm{e}^{-\frac{1}{2}w}$. F 和 W 的联合密度 $(\mathrm{d}v=pw\mathrm{d}f/m)$ 是

$$\begin{aligned}&\frac{\mathrm{e}^{-\frac{1}{2}\tau^2}}{2^{\frac{1}{2}(p+m)}\Gamma(\frac{1}{2}m)}\mathrm{e}^{-\frac{1}{2}w(1+pf/m)} \\ &\quad\cdot\frac{p}{m}\sum_{\beta=0}^{\infty}\left(\frac{\tau^2}{4}\right)^{\beta}\frac{1}{\beta!\Gamma(\frac{1}{2}p+\beta)}\left(\frac{pf}{m}\right)^{\frac{1}{2}p+\beta-1}w^{\frac{1}{2}(p+m)+\beta-1}.\end{aligned} \tag{1}$$

通过对 (1) 关于 w 从 0 到 ∞ 积分, 得到边缘密度

$$\frac{p\mathrm{e}^{-\frac{1}{2}\tau^2}}{m\Gamma(\frac{1}{2}m)}\sum_{\beta=0}^{\infty}\frac{(\tau^2/2)^{\beta}(pf/m)^{\frac{1}{2}p+\beta-1}\Gamma\left[\frac{1}{2}(p+m)+\beta\right]}{\beta!\Gamma(\frac{1}{2}p+\beta)(1+pf/m)^{\frac{1}{2}(p+m)+\beta}}. \tag{2}$$

定理 5.4.1 如果 V 服从自由度为 p, 非中心参数为 τ^2 的非中心 χ^2 分布, W 服从一个独立的自由度为 m 的 χ^2 分布, 则 $F=(V/p)/(W/m)$ 有密度 (2).

密度 (2) 是非中心 F 分布的密度.

如果 $T^2=N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)$ 是基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本, 则 $(T^2/n)(N-p)/p$ 服从自由度为 p 和 $N-p$, 非中心参数为 $N(\boldsymbol{\mu}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}_0)=\tau^2$ 的非中心 F 分布. 由 (2), 我们求得 T^2 的密度是

$$\frac{e^{-\frac{1}{2}\tau^2}}{(N-1)\Gamma[\frac{1}{2}(N-p)]}\sum_{\beta=0}^{\infty}\frac{(\tau^2/2)^\beta[t^2/(N-1)]^{\frac{1}{2}p+\beta-1}\Gamma(\frac{1}{2}N+\beta)}{\beta!\Gamma(\frac{1}{2}p+\beta)[1+t^2/(N-1)]^{\frac{1}{2}N+\beta}} \tag{3}$$
$$=\frac{\Gamma(\frac{1}{2}N)}{(N-1)\Gamma[\frac{1}{2}(N-p)]\Gamma(\frac{1}{2}p)}\left(\frac{t^2}{N-1}\right)^{\frac{1}{2}p-1}\left(1+\frac{t^2}{N-1}\right)^{-\frac{1}{2}N}$$
$$\cdot e^{-\frac{1}{2}\tau^2}{}_1F_1\left(\frac{1}{2}N;\frac{1}{2}p;\frac{\tau^2t^2}{2(N-1)}\right).$$

其中

$$_1F_1(a;b;x)=\sum_{\beta=0}^{\infty}\frac{\Gamma(a+\beta)\Gamma(b)x^\beta}{\Gamma(a)\Gamma(b+\beta)\beta!}. \tag{4}$$

密度 (3) 是非中心 T^2 分布的密度.

Tang (1938) 对不同 τ^2 值, 显著性水平 0.05 和 0.01, 列出了接受原假设的概率 (即第二类错误的概率). 他的自由度 f_1 是我们的 $p[1(1)8]$, 他的 f_2 是我们的 $n-p+1[2,4(1)30,60,\infty]$, 他的非中心参数 ϕ 与我们的 τ^2 有关系

$$\phi=\frac{\tau}{\sqrt{p+1}} \tag{5}$$

$[1(\frac{1}{2})3(1)8]$. 他的分位数的伴随表格是针对 $T^2/(T^2+N-1)$ 的.

例如, 假设 $p=4, n-p+1=20$, 考虑以显著性水平 1% 检验原假设 $\boldsymbol{\mu}=\mathbf{0}$. 我们想知道, 比如说当 $\phi=2.5$ $(\tau^2=31.25)$ 时, 我们接受原假设的概率. 它是 0.227. 如果我们认为当 $N,\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 满足 $\tau^2=31.25$ 时接受原假设带来的弊处小于当原假设为真时拒绝它的弊处, 则我们认为有理由如假定的进行检验. 但是, 如果这两类错误的弊处都一样, 那么有理由降低第二类错误的概率. 从而, 如果我们取显著性水平 5%, 第二类错误的概率 (对 $\phi=2.5$) 只是 0.043.

Lehmer (1944) 对给定显著性水平和第二类错误的概率, 计算出了 ϕ 的表. 这里可以用这些表查看当 $\boldsymbol{\mu}\neq 0$ 时, 使得接受原假设的概率足够小时所需要的 τ^2 值. 比如, 如果对于给定的 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$, 在某样本基础上, 我们想拒绝假设 $\boldsymbol{\mu}=0$, 可以选择 N, 使得 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}=\tau^2$ 充分大. 当然, 困难的是, 当我们想让拒绝概率达到某个特定值时, 我们通常不知道 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ (从而 τ^2) 精确的值.

当原假设不真时, T^2 的分布可以通过不同的方法得到, 见 Hsu (1938) 及 Bose and Roy (1938).

5.5　协方差阵不等时的两样本问题

如果协方差阵不相等, 均值向量相等的 T^2 检验在原假设下的拒绝概率依赖它们的协方差阵. 如果两个协方差阵的差别很小或样本容量足够大, 则不存在实际影响. 但是, 如果协方差阵差别太大且 (或者) 样本容量很小, 则名义上的显著

性水平会受到干扰. 因此, 对于指定的显著性水平, 我们提出一个新的方法. 设 $\{\boldsymbol{x}_\alpha^{(i)}\}$ $(\alpha = 1,\cdots,N_i)$ 是来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_i)$ $(i=1,2)$ 的样本. 我们想检验假设 $H:\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}$. 第一个样本的均值 $\bar{\boldsymbol{x}}^{(1)}$ 服从正态分布, 其期望是

$$\mathrm{E}(\bar{\boldsymbol{x}}^{(1)}) = \boldsymbol{\mu}^{(1)}, \tag{1}$$

协方差阵是

$$\mathrm{E}(\bar{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)})(\bar{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)})' = \frac{1}{N_1}\boldsymbol{\Sigma}_1. \tag{2}$$

类似地, 第二个样本的均值 $\bar{\boldsymbol{x}}^{(2)}$ 服从正态分布, 其均值是

$$\mathrm{E}(\bar{\boldsymbol{x}}^{(2)}) = \boldsymbol{\mu}^{(2)}, \tag{3}$$

协方差阵是

$$\mathrm{E}(\bar{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})(\bar{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})' = \frac{1}{N_2}\boldsymbol{\Sigma}_2. \tag{4}$$

因此 $\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)}$ 有均值 $\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}$, 协方差阵 $(1/N_1)\boldsymbol{\Sigma}_1+(1/N_2)\boldsymbol{\Sigma}_2$. 但是我们不能用 5.2 节的方法, 因为

$$\sum_{\alpha}^{N_1}(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})' + \sum_{\alpha=1}^{N_2}(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})' \tag{5}$$

不服从协方差阵为 $(1/N_1)\boldsymbol{\Sigma}_1+(1/N_2)\boldsymbol{\Sigma}_2$ 的倍数的 Wishart 分布.

如果 $N_1=N_2=N$, 我们可以通过显然的方法利用 T^2 检验. 设 $\boldsymbol{y}_\alpha=\boldsymbol{x}_\alpha^{(1)}-\boldsymbol{x}_\alpha^{(2)}$ (假定两组样本的观测个数与观测本身独立). 则 $\boldsymbol{y}_\alpha$ 服从均值为 $\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}$、协方差阵为 $\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2$ 的正态分布, 且 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N$ 独立. 设 $\bar{\boldsymbol{y}}=(1/N)\sum_{\alpha=1}^N\boldsymbol{y}_\alpha=\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)}$, 定义 $\boldsymbol{S}$ 为

$$\begin{aligned}(N-1)\boldsymbol{S} &= \sum_{\alpha=1}^{N}(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})' \\ &= \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^{(1)}-\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(1)}+\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(1)}-\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(1)}+\bar{\boldsymbol{x}}^{(2)})'.\end{aligned} \tag{6}$$

则

$$T^2 = N\bar{\boldsymbol{y}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{y}} \tag{7}$$

可用来检验假设 $\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}=\boldsymbol{0}$, 它服从自由度为 $N-1$ 的 T^2 分布. 需要注意的是, 如果我们知道 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$, 则我们将用一个自由度为 $2N-2$ 的 T^2 统计量, 因此我们损失了 $N-1$ 个自由度去构造一个与两个协方差阵独立的检验. 如果 $N_1=N_2=50$, 如 5.3.4 节的例子, 则 $T^2_{4,49}(0.01)=15.93$, 而 $T^2_{4,98}(0.01)=14.52$.

现在让我们考虑 $N_1\neq N_2$ 的情况. 为了方便, 设 $N_1<N_2$. 则我们定义

$$\boldsymbol{y}_\alpha=\boldsymbol{x}_\alpha^{(1)}-\sqrt{\frac{N_1}{N_2}}\boldsymbol{x}_\alpha^{(2)}+\frac{1}{\sqrt{N_1N_2}}\sum_{\beta=1}^{N_1}\boldsymbol{x}_\beta^{(2)}-\frac{1}{N_2}\sum_{\gamma=1}^{N_2}\boldsymbol{x}_\gamma^{(2)}, \qquad \alpha=1,\cdots,N_1. \tag{8}$$

$\boldsymbol{y}_\alpha$ 的期望值是

$$\mathrm{E}(\boldsymbol{y}_\alpha) = \boldsymbol{\mu}^{(1)} - \sqrt{\frac{N_1}{N_2}}\boldsymbol{\mu}^{(2)} + \frac{N_1}{\sqrt{N_1 N_2}}\boldsymbol{\mu}^{(2)} - \frac{N_2}{N_2}\boldsymbol{\mu}^{(2)} = \boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}. \tag{9}$$

$\boldsymbol{y}_\alpha$ 和 $\boldsymbol{y}_\beta$ 的协方差阵是

$$\mathrm{E}\left[(\boldsymbol{y}_\alpha - \mathrm{E}(\boldsymbol{y}_\alpha))(\boldsymbol{y}_\beta - \mathrm{E}(\boldsymbol{y}_\beta))'\right] = \delta_{\alpha\beta}\left(\boldsymbol{\Sigma}_1 + \frac{N_1}{N_2}\boldsymbol{\Sigma}_2\right). \tag{10}$$

因此检验 $\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)} = \mathbf{0}$ 的一个合适的统计量是

$$T^2 = N_1 \bar{\boldsymbol{y}}' \boldsymbol{S}^{-1} \bar{\boldsymbol{y}}, \tag{11}$$

其中

$$\bar{\boldsymbol{y}} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1} \boldsymbol{y}_\alpha = \bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}; \tag{12}$$

$$(N_1 - 1)\boldsymbol{S} = \sum_{\alpha=1}^{N_1}(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})' = \sum_{\alpha=1}^{N_1}(\boldsymbol{u}_\alpha - \bar{\boldsymbol{u}})(\boldsymbol{u}_\alpha - \bar{\boldsymbol{u}})', \tag{13}$$

其中 $\bar{\boldsymbol{u}} = (1/N_1)\sum_{\alpha=1}^{N_1} \boldsymbol{u}_\alpha$ 和 $\boldsymbol{u}_\alpha = \boldsymbol{x}_\alpha^{(1)} - \sqrt{N_1/N_2}\boldsymbol{x}_\alpha^{(2)}, \alpha = 1, \cdots, N_1$. 这个统计量服从自由度为 $N_1 - 1$ 的 T^2 分布.

这个方法是 Scheffé (1943) 对单变量情况提出的. Scheffé 证明了在单变量情况下, 这个方法可由 t 分布得到最短的置信区间. 这个方法的优点是用到了 $\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}$, 这个统计量和 $\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ 相关最大. 在估计协方差阵时观测的牺牲并不是那么重要. Bennett (1951) 给出了这个方法在多元情况的推广.

这个方法可以用在更一般的情况下. 设 $\{\boldsymbol{x}_\alpha^{(i)}\}$ $(\alpha = 1, \cdots, N_i, i = 1, \cdots, q)$ 是分别来自 $N(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}_i)$ $(i = 1, \cdots, q)$ 的样本. 考虑检验假设

$$H: \sum_{i=1}^{q} \beta_i \boldsymbol{\mu}^{(i)} = \boldsymbol{\mu}, \tag{14}$$

其中 $\beta_1, \cdots, \beta_q$ 是给定的标量且 $\boldsymbol{\mu}$ 是给定的向量. 如果 N_i 不等, 取 N_1 为最小的. 设

$$\boldsymbol{y}_\alpha = \beta_1 \boldsymbol{x}_\alpha^{(1)} + \sum_{i=2}^{q} \beta_i \sqrt{\frac{N_1}{N_i}}\left(\boldsymbol{x}_\alpha^{(i)} - \frac{1}{N_1}\sum_{\beta=1}^{N_1}\boldsymbol{x}_\beta^{(i)} + \frac{1}{\sqrt{N_1 N_i}}\sum_{\gamma=1}^{N_i}\boldsymbol{x}_\gamma^{(i)}\right). \tag{15}$$

则 $\mathrm{E}(\boldsymbol{y}_\alpha) = \sum_{i=1}^{q} \beta_i \boldsymbol{\mu}^{(i)}$, 且

$$\mathrm{E}\left[(\boldsymbol{y}_\alpha - \mathrm{E}(\boldsymbol{y}_\alpha))(\boldsymbol{y}_\beta - \mathrm{E}(\boldsymbol{y}_\beta))'\right] = \delta_{\alpha\beta}\left(\sum_{i=1}^{q}\frac{\beta_i^2 N_1}{N_i}\boldsymbol{\Sigma}_i\right). \tag{16}$$

设 $\bar{\boldsymbol{y}}$ 和 $\boldsymbol{S}$ 定义为

$$\bar{\boldsymbol{y}} = \frac{1}{N_1}\sum_{\alpha=1}^{N_1}\boldsymbol{y}_\alpha = \sum_{i=1}^{q}\beta_i \bar{\boldsymbol{x}}^{(i)}, \qquad \bar{\boldsymbol{x}}^{(i)} = \frac{1}{N_i}\sum_{\beta=1}^{N_i}\boldsymbol{x}_\beta^{(i)}, \tag{17}$$

$$(N_1 - 1)\boldsymbol{S} = \sum_{\alpha=1}^{N_1}(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})'. \tag{18}$$

则

$$T^2 = N_1(\bar{\boldsymbol{y}} - \boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}) \tag{19}$$

对检验 H 是合适的, 且当假设为真时, 这个统计量服从自由度为 $N_1 - 1$ 的 p 维 T^2 分布. 如果我们设 $\boldsymbol{u}_\alpha = \sum_{i=1}^q \beta_i \sqrt{N_1/N_i}\boldsymbol{x}_\alpha^{(i)}, \alpha = 1, \cdots, N_1$, 则 $\boldsymbol{S}$ 可以定义为

$$(N_1 - 1)\boldsymbol{S} = \sum_{\alpha=1}^{N_1}(\boldsymbol{u}_\alpha - \bar{\boldsymbol{u}})(\boldsymbol{u}_\alpha - \bar{\boldsymbol{u}})'. \tag{20}$$

另一个可用这种方法处理的问题是检验两个子向量有相同均值的假设. 设 $\boldsymbol{x} = (\boldsymbol{x}^{(1)\prime}, \boldsymbol{x}^{(2)\prime})'$ 服从均值为 $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)\prime}, \boldsymbol{\mu}^{(2)\prime})$ 和协方差阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{21}$$

的正态分布. 我们假定 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 分别有 q 个分量. 则 $\boldsymbol{y} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}$ 服从均值为 $\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}$ 和协方差阵为 $\boldsymbol{\Sigma}_{\boldsymbol{y}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{21} - \boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{22}$ 的正态分布. 为了检验假设 $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$, 我们用 T^2 统计量 $N\bar{\boldsymbol{y}}'\boldsymbol{S}_{\boldsymbol{y}}^{-1}\bar{\boldsymbol{y}}$, 其中样本均值向量和协方差阵的划分类似于 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的划分形式.

5.6 T^2 检验的一些最优性质

5.6.1 最优不变检验

本节我们将说明 T^2 检验在某特定检验类中是最优的, 并简单地介绍一下这些结果的证明.

基于来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 N 个观测 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, 检验假设 $\boldsymbol{\mu} = \boldsymbol{0}$. 首先, 我们考虑基于统计量 $\boldsymbol{A} = \sum(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'$ 和 $\bar{\boldsymbol{x}}$, 关于变换 $\boldsymbol{A}^* = \boldsymbol{CAC}'$ 和 $\bar{\boldsymbol{x}}^* = \boldsymbol{C}\bar{\boldsymbol{x}}$ 不变的检验类, 其中 $\boldsymbol{C}$ 非奇异. 变换 $\boldsymbol{x}_\alpha^* = \boldsymbol{C}\boldsymbol{x}_\alpha$ 保持问题的不变性, 即以 $\boldsymbol{x}_\alpha^*$ 的角度, 给定来自一个多元正态总体的 N 个观测 $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_N^*$, 我们检验假设 $\mathrm{E}(\boldsymbol{x}_\alpha^*) = \boldsymbol{0}$. 看起来我们需要一个同样在这些变换下不变的解, 即我们要寻找一个临界区域, 它在非奇异线性变换下不变. (在不同坐标系下区域的定义是一样的.)

定理 5.6.1 *给定来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的样本 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, 在所有基于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A} = \boldsymbol{\Sigma}(x_\alpha - \bar{x})(x_\alpha - \bar{x})'$ 的关于变换 $\bar{\boldsymbol{x}}^* = \boldsymbol{C}\bar{\boldsymbol{x}}, \boldsymbol{A}^* = \boldsymbol{CAC}'$ ($\boldsymbol{C}$ 非奇异) 不变的 $\boldsymbol{\mu} = \boldsymbol{0}$ 检验中, T^2 检验是一致最优的.*

证明 首先, 我们在 5.2.1 节已经知道, 任何基于 T^2 的检验都是不变的. 其次, 这个函数本质上是唯一不变的, 因为如果 $f(\bar{\boldsymbol{x}}, \boldsymbol{A})$ 是不变的, 则 $f(\bar{\boldsymbol{x}}, \boldsymbol{A}) = f(\bar{\boldsymbol{x}}^*, \boldsymbol{I})$, 其中, $\bar{\boldsymbol{x}}^*$ 只有第一个坐标非零且为 $\sqrt{\bar{\boldsymbol{x}}'\boldsymbol{A}^{-1}\bar{\boldsymbol{x}}}$. (存在一个矩阵 $\boldsymbol{C}$, 满足 $\boldsymbol{C}\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}^*$ 和 $\boldsymbol{CAC}' = \boldsymbol{I}$.) 因此, $f(\bar{\boldsymbol{x}}, \boldsymbol{A})$ 只依赖 $\bar{\boldsymbol{x}}'\boldsymbol{A}^{-1}\bar{\boldsymbol{x}}$. 从而一个不变检验一定基于 $\bar{\boldsymbol{x}}'\boldsymbol{A}^{-1}\bar{\boldsymbol{x}}$. 然后, 我们可以对 T^2 的分布 [5.4 节的 (3)] 应用奈曼-皮尔逊基本引理以求基于 T^2

及其简单备择 $\tau^2=N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的一致最优检验. $\tau^2=0$ 的一致最优检验是基于 5.4 节的 (3) 和 $\tau^2=0$ 时的 (3) 的比. 临界区域是

$$c<\mathrm{e}^{-\frac{1}{2}\tau^2}\sum_{\alpha=0}^{\infty}\frac{(\tau^2/2)^{\alpha}(t^2/n)^{\frac{1}{2}p+\alpha-1}(1+t^2/n)^{-\frac{1}{2}(n+1)-\alpha}\Gamma[\frac{1}{2}(n+1)+\alpha]}{\alpha!\Gamma(\frac{1}{2}p+\alpha)} \tag{1}$$

$$\Bigg/\frac{(t^2/n)^{\frac{1}{2}p-1}(1+t^2/n)^{-\frac{1}{2}(n+1)}\Gamma[\frac{1}{2}(n+1)]}{\Gamma(\frac{1}{2}p)}$$

$$=\frac{\Gamma(\frac{1}{2}p)}{\Gamma[\frac{1}{2}(n+1)]}\mathrm{e}^{-\frac{1}{2}\tau^2}\sum_{\alpha=0}^{\infty}\frac{(\tau^2/2)^{\alpha}\Gamma[\frac{1}{2}(n+1)+\alpha]}{\alpha!\Gamma(\frac{1}{2}p+\alpha)}\left(\frac{t^2/n}{1+t^2/n}\right)^{\alpha}.$$

(1) 的右端项是 $(t^2/n)/(1+t^2/n)$ 的严格增函数, 从而也是 t^2 的严格增函数. 因此, 这个不等式对于适当的 k 值等价于 $t^2>k$. 因为不依赖备择 τ^2, 所以这个检验是一致最优不变的. ■

定义 5.6.1　检验函数 $\psi(\bar{\boldsymbol{x}},\boldsymbol{A})$ 是值域为 0 到 1 的一个函数, 它满足当 $\boldsymbol{\mu}=\boldsymbol{0}$ 时, $\mathrm{E}[\psi(\bar{\boldsymbol{x}},\boldsymbol{A})]=\varepsilon$, 即显著性水平.

一个随机检验就是当 $\bar{\boldsymbol{x}}=\boldsymbol{x}$ 和 $\boldsymbol{A}=\boldsymbol{B}$ 时, 以概率 $\psi(\boldsymbol{x},\boldsymbol{B})$ 拒绝假设. 一个非随机检验就是 $\psi(\bar{\boldsymbol{x}},\boldsymbol{A})$ 只取值 0 和 1. 利用适当的检验函数的奈曼–皮尔逊引理形式, 我们得到下面的推论.

推论 5.6.1　在一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的基础上, 所有基于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}$ 的关于变换 $\bar{\boldsymbol{x}}^*=\boldsymbol{C}\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}^*=\boldsymbol{CAC}'$ ($\boldsymbol{C}$ 非奇异) 不变的随机检验中, T^2 检验是一致最优的.

定理 5.6.2　在一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的基础上, 所有关于变换 $\boldsymbol{x}_\alpha^*=\boldsymbol{C}\boldsymbol{x}_\alpha$ 不变的 $\boldsymbol{\mu}=\boldsymbol{0}$ 的检验中, T^2 检验是一个一致最优检验, 即 T^2 检验至少也和其他不变检验的功效相同.

证明　设 $\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 是一个不变检验的检验函数. 则

$$\mathrm{E}[\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)]=\mathrm{E}_{\bar{\boldsymbol{x}},\boldsymbol{A}}\{\mathrm{E}[\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)|\bar{\boldsymbol{x}},\boldsymbol{A}]\}. \tag{2}$$

因为 $\bar{\boldsymbol{x}},\boldsymbol{A}$ 是 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的充分统计量, 所以期望 $\mathrm{E}[\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)|\bar{\boldsymbol{x}},\boldsymbol{A}]$ 只依赖 $\bar{\boldsymbol{x}},\boldsymbol{A}$. 它是不变的, 且和 $\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 功效相同. 因此, 每一个这样较大类中的检验都可以由一个较小类 (只依赖 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}$) 的检验代替, 它们的功效相同. 再由推论 5.6.1, 定理得证. ■

定理 5.6.3　给定来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 在所有基于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}=\sum(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$, 且功效只依赖 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的检验中, T^2 检验是一致最优的.

证明　我们希望通过建立功效只依赖 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的检验类与不变检验类的一致性, 将这个定理简化为定理 5.6.1. 我们需要下面的定义.

定义 5.6.2　检验 $\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 称为**几乎不变的**, 如果对所有 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 有

$$\psi(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)=\psi(\boldsymbol{C}\boldsymbol{x}_1,\cdots,\boldsymbol{C}\boldsymbol{x}_N), \tag{3}$$

除了 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的一个勒贝格零测集, 外 (这个例外集可以依赖 $\boldsymbol{C}$).

显然, 如果我们把不变检验的定义推广为, (3) 在除了一个固定的 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的勒贝格零测集外成立 (这个集合不依赖 $\boldsymbol{C}$), 则定理 5.6.1 和定理 5.6.2 成立. Hunt and Stein [Lehmann (1959)] 指出在我们的问题里, 几乎不变性意味着不变性 (从广义的角度).

现在, 我们想讨论, 如果 $\psi(\bar{x},\boldsymbol{A})$ 的功效只依赖 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, 那么它是几乎不变的. 因为 $\psi(\bar{x},\boldsymbol{A})$ 的功效只依赖 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, 其功效为

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\psi(\bar{\boldsymbol{x}},\boldsymbol{A})&\equiv \mathrm{E}_{\boldsymbol{C}^{-1}\boldsymbol{\mu},\ \boldsymbol{C}^{-1}\boldsymbol{\Sigma}(\boldsymbol{C}^{-1})'}\psi(\bar{\boldsymbol{x}},\boldsymbol{A})\\&\equiv \mathrm{E}_{\boldsymbol{\mu},\ \boldsymbol{\Sigma}}\psi(\boldsymbol{C}\bar{\boldsymbol{x}},\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}').\end{aligned} \tag{4}$$

(4) 的第二项和第三项只是在积分表达形式上不同. 因此

$$\mathrm{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\left[\psi(\bar{\boldsymbol{x}},\boldsymbol{A})-\psi(\boldsymbol{C}\bar{\boldsymbol{x}},\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}')\right]\equiv 0, \tag{5}$$

对 $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, 完全相同. 因为 $\bar{\boldsymbol{x}},\boldsymbol{A}$ 是 $\boldsymbol{\mu},\boldsymbol{\Sigma}$ 的一个充分统计量集合 (定理 3.4.2), 所以 $f(\bar{\boldsymbol{x}},\boldsymbol{A})=\psi(\bar{\boldsymbol{x}},\boldsymbol{A})-\psi(\boldsymbol{C}\bar{\boldsymbol{x}},\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}')=0$ 几乎处处成立. 定理 5.6.3 得证. ■

由于定理 5.6.2 由定理 5.6.1 而得, 从而下面的定理可由定理 5.6.3 得证.

定理 5.6.4 *在一组来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的基础上, 所有功效只依赖 $N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的 $\boldsymbol{\mu}=\boldsymbol{0}$ 的检验中, T^2 检验是一致最优检验.*

定理 5.6.4 首先由 Simaika (1941) 得出. 本节给出的结果和证明来自 Lehmann (1959). Hsu (1945) 证明了 T^2 检验的一个关于 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 上平均功效的最优性质.

5.6.2 容许检验

我们现在来考虑 T^2 检验相比所有可能的检验是否是一个好的检验 (前面一节的对比是限制在不变检验类中的). 得到的主要结果是 T^2 检验是所有检验类中容许的, 即没有其他方法比它更好.

定义 5.6.3 *一个关于原假设 $H_0:\omega\in\Omega_0$ 及其备择 $\omega\in\Omega_1$ (与 Ω_0 不相交) 的检验 T^* 是容许的, 如果不存在其他检验 T, 满足*

$$\Pr\{拒绝H_0|T,\omega\}\leqslant \Pr\{拒绝H_0|T^*,\omega\},\qquad \omega\in\Omega_0, \tag{6}$$

$$\Pr\{拒绝H_0|T,\omega\}\geqslant \Pr\{拒绝H_0|T^*,\omega\},\qquad \omega\in\Omega_1. \tag{7}$$

其中, 不等号至少对某一个 ω 严格成立.

T^2 检验的容许性可由 Stein (1956a) 的一个定理应用到任何指数分布族得到.

一个指数分布族$(\mathscr{Y},\mathscr{B},m,\Omega,P)$ 由以下几个元素构成: 一个有限维欧式空间 $\mathscr{Y}$; 一个 $\mathscr{Y}$ 的所有寻常博雷尔集的 σ 代数 $\mathscr{B}$ 上的测度 m; 伴随空间 $\mathscr{Y}'$ ($\mathscr{Y}$ 上所有实值线性函数的线性空间) 的一个满足

$$\psi(\boldsymbol{\omega})=\int_{\mathscr{Y}}\mathrm{e}^{\boldsymbol{\omega}'\boldsymbol{y}}\mathrm{d}m(\boldsymbol{y})<\infty,\quad \boldsymbol{\omega}\in\Omega \tag{8}$$

的子集 Ω; 从 Ω 到 $\mathscr{B}$ 上概率测度集合的函数 P,

$$P_{\boldsymbol{\omega}}(A)=\frac{1}{\psi(\boldsymbol{\omega})}\int_A \mathrm{e}^{\boldsymbol{\omega}'\boldsymbol{y}}\mathrm{d}m(\boldsymbol{y}),\qquad A\in\mathscr{B}.$$

正态分布族 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 是一类指数分布族, 因为它的密度可以写成

$$n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{e}^{(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1})\boldsymbol{x}+\mathrm{tr}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1})\boldsymbol{x}\boldsymbol{x}'}. \tag{9}$$

我们把 $\mathscr{X}$ 映射到 $\mathscr{Y}$, 向量 $\boldsymbol{y}=(\boldsymbol{y}^{(1)\prime},\boldsymbol{y}^{(2)\prime})'$ 由 $\boldsymbol{y}^{(1)}=\boldsymbol{x}$ 和 $\boldsymbol{y}^{(2)}=(x_1^2,2x_1x_2,\cdots,2x_1x_p,x_2^2,\cdots,x_p^2)'$ 组成. 向量 $\boldsymbol{\omega}=(\boldsymbol{\omega}^{(1)\prime},\boldsymbol{\omega}^{(2)\prime})'$ 由 $\boldsymbol{\omega}^{(1)}=\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 和 $\boldsymbol{\omega}^{(2)}=-\frac{1}{2}(\sigma^{11},\sigma^{12},\cdots,\sigma^{1p},\sigma^{22},\cdots,\sigma^{pp})'$ 组成的, 其中 $(\sigma^{ij})=\boldsymbol{\Sigma}^{-1}$, 参数变换一一对应. 集合 $A\in\mathscr{B}$ 的测度 $m(A)$ 是 $\boldsymbol{x}$ 的集合映射到集合 A 的普通勒贝格测度. (注意 $\mathscr{Y}$ 中的概率测度不是由一个密度定义的.)

定理 5.6.5 (Stein)　*设 $(\mathscr{Y},\mathscr{B},m,\Omega,P)$ 是一个指数族, Ω_0 是 Ω 的一个非空真子集. (i) 设 A 是 $\mathscr{Y}$ 的一个闭凸子集. (ii) 假定对每个向量 $\boldsymbol{\omega}\in\mathscr{Y}'$ 和使得 $\{\boldsymbol{y}|\boldsymbol{\omega}'\boldsymbol{y}>c\}$ 与 A 不相交的实数 c, 存在 $\boldsymbol{\omega}_1\in\Omega$, 使得对任意大的 λ, 向量 $\boldsymbol{\omega}_1+\lambda\boldsymbol{\omega}\in\Omega-\Omega_0$. 则对于假设检验 $\boldsymbol{\omega}\in\Omega_0$ 及其备择 $\boldsymbol{\omega}\in\Omega-\Omega_0$, 接受区域为 A 的检验是容许的.*

图 5.2 在 $\mathscr{Y}$ 空间和 Ω 集合中, 同时说明了定理中的条件.

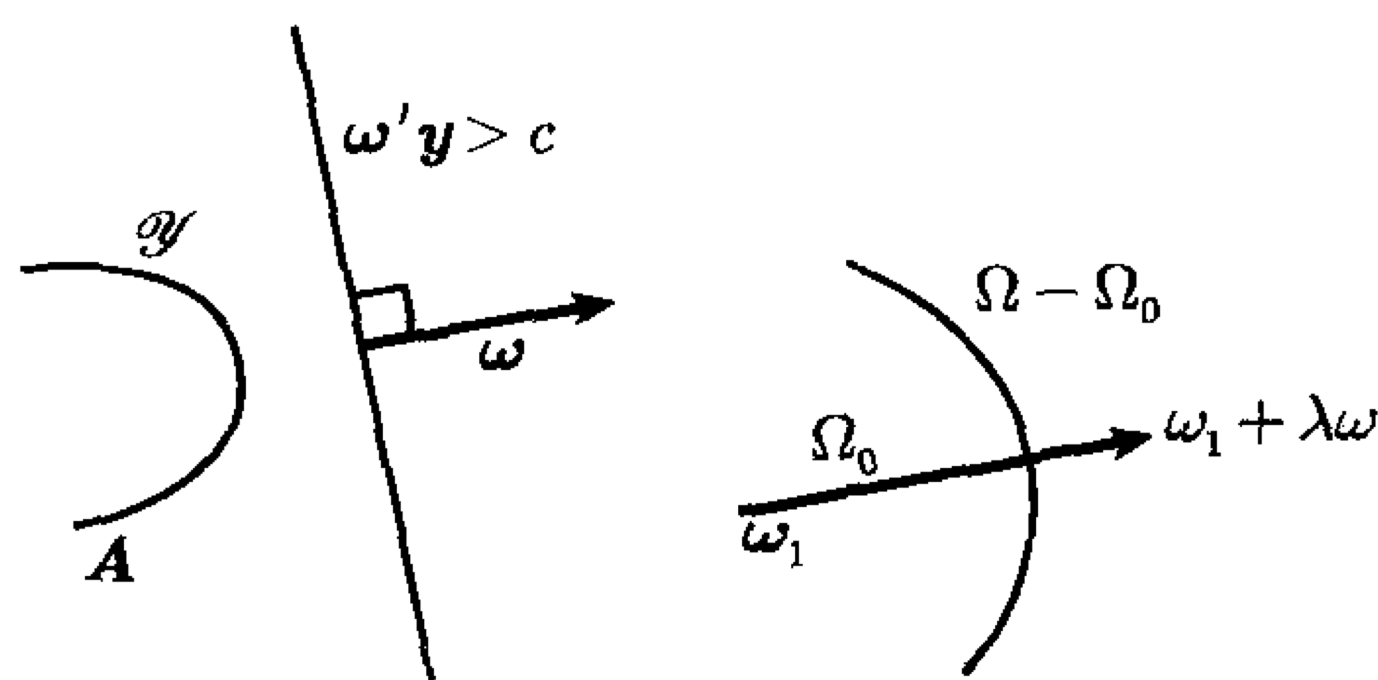


图　5.2

证明　接受区域为 A 的检验的检验函数是 $\phi_A(\boldsymbol{y})=0$ $(\boldsymbol{y}\in A)$ 和 $\phi_A(\boldsymbol{y})=1$ $(\boldsymbol{y}\notin A)$. 假定 $\phi(\boldsymbol{y})$ 是一个更优检验的检验函数, 即

$$\int\phi(\boldsymbol{y})\mathrm{d}P_{\boldsymbol{\omega}}(\boldsymbol{y})\leqslant\int\phi_A(\boldsymbol{y})\mathrm{d}P_{\boldsymbol{\omega}}(\boldsymbol{y}),\quad \boldsymbol{\omega}\in\Omega_0, \tag{10}$$

$$\int\phi(\boldsymbol{y})\mathrm{d}P_{\boldsymbol{\omega}}(\boldsymbol{y})\geqslant\int\phi_A(\boldsymbol{y})\mathrm{d}P_{\boldsymbol{\omega}}(\boldsymbol{y}),\quad \boldsymbol{\omega}\in\Omega-\Omega_0, \tag{11}$$

且对某一个 $\boldsymbol{\omega}$, 不等式严格成立, 我们来证明这个假定将产生矛盾. 设 $B=\{\boldsymbol{y}|\phi(\boldsymbol{y})<1\}$. (如果对比的检验是非随机化的, 则 B 是其接受区域.) 则

$$\{y|\phi_A(y)-\phi(y)>0\}=\bar{A}\cap B, \tag{12}$$

其中 $\bar{A}$ 是 A 的补集. (12) 的 m 测度为正, 否则 $\phi_A(y)=\phi(y)$ 几乎处处成立, 则 (10) 和 (11) 对所有 ω, 等号成立. 因为 A 是凸集, 则存在一个 ω 和一个 c, 使得 $\bar{A}\cap B$ 和 $\{y|\omega'y>c\}$ 的交集有正 m 测度. (因为 A 是闭的, 则 $\bar{A}$ 是开的, 从而它可以由可数个开球覆盖, 比如, 有理数半径和中心坐标为有理数的球面. 因为存在一个超平面将 A 和每一个球面分离, 所以存在可数个与 A 不交的开半空间 H_j, 它覆盖了 $\bar{A}$. 则至少有一个半空间和 $\bar{A}\cap B$ 有 m 测度为正的交集.) 由假设, 存在 $\omega_1\in\Omega$ 和任意大的 λ, 使得

$$\omega_\lambda=\omega_1+\lambda\omega\in\Omega-\Omega_0. \tag{13}$$

则

$$\begin{aligned}
&\int[\phi_A(y)-\phi(y)]\mathrm{d}P_{\omega_\lambda}(y) \qquad (14)\\
=&\frac{1}{\psi(\omega_\lambda)}\int[\phi_A(y)-\phi(y)]\mathrm{e}^{\omega_\lambda' y}\mathrm{d}m(y)\\
=&\frac{\psi(\omega_1)}{\psi(\omega_\lambda)}\int[\phi_A(y)-\phi(y)]\mathrm{e}^{\lambda\omega' y}\mathrm{d}P_{\omega_1}(y)\\
=&\frac{\psi(\omega_1)}{\psi(\omega_\lambda)}\mathrm{e}^{\lambda c}\int[\phi_A(y)-\phi(y)]\mathrm{e}^{\lambda(\omega' y-c)}\mathrm{d}P_{\omega_1}(y)\\
=&\frac{\psi(\omega_1)}{\psi(\omega_\lambda)}\mathrm{e}^{\lambda c}\left\{\int_{\omega' y>c}[\phi_A(y)-\phi(y)]\mathrm{e}^{\lambda(\omega' y-c)}\mathrm{d}P_{\omega_1}(y)\right.\\
&\left.+\int_{\omega' y\leqslant c}[\phi_A(y)-\phi(y)]\mathrm{e}^{\lambda(\omega' y-c)}\mathrm{d}P_{\omega_1}(y)\right\}.
\end{aligned}$$

对于 $\omega'y>c$, 有 $\phi_A(y)=1$ 和 $\phi_A(y)-\phi(y)\geqslant 0$, 并且 $\{y|\phi_A(y)-\phi(y)>0\}$ 有正测度; 因此, 花括号里的第一个积分当 $\lambda\to\infty$ 时, 趋于 ∞. 第二个积分是有界的, 因为积分的界为 1, 从而最后一个表达式当 λ 充分大时是正的. 这与 (11) 矛盾. ■

这个证明是 Stein (1956a) 给出的. 它是 Birnhaum (1955) 的一个定理推广.

推论 5.6.2 如果除了 A 不必为闭集, 但是 A 的边界的 m 测度为 0 外, 定理 5.6.5 的条件成立, 则定理 5.6.5 的结论成立.

证明 A 的闭包是凸的 (习题 5.18), 且对所有 $\omega\in\Omega$, 检验的接受区域等于 A 的闭包而且只在一个零概率集上不同于 A. 从而,

$$\begin{aligned}
A\cap\{y|\omega'y>c\}=\varnothing&\Rightarrow A\subset\{y|\omega'y\leqslant c\} \qquad (15)\\
&\Rightarrow A \text{ 的闭包 } \subset\{y|\omega'y\leqslant c\}.
\end{aligned}$$

则用 A 的闭包代替 A, 定理 5.6.5 成立. ■

定理 5.6.6 基于来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测 $x_1,\cdots,x_N$, 对检验假设 $\boldsymbol{\mu}=\mathbf{0}$, Hotelling T^2 检验是容许的.

证明 为了应用定理 5.6.5, 我们把观测的分布写成指数族形式. 由定理 3.3.1 和定理 3.3.2, 我们可以变换 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 为 $\boldsymbol{z}_\alpha=\sum_{\beta=1}^N c_{\alpha\beta}\boldsymbol{x}_\beta$, 其中 $(c_{\alpha\beta})$ 是正交的且 $\boldsymbol{z}_N=\sqrt{N}\bar{\boldsymbol{x}}$. 则 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N$ 的密度 (关于勒贝格测度) 是

$$\frac{\mathrm{e}^{-\frac{1}{2}N\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{\frac{1}{2}N}}\exp\left[\sqrt{N}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_N+\operatorname{tr}(-\frac{1}{2}\boldsymbol{\Sigma}^{-1})\sum_{\alpha=1}^N\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'\right]. \tag{16}$$

向量 $\boldsymbol{y}=(\boldsymbol{y}^{(1)},\boldsymbol{y}^{(2)\prime})'$ 由 $\boldsymbol{y}^{(1)}=\boldsymbol{z}_N\ (=\sqrt{N}\bar{\boldsymbol{x}})$ 和 $\boldsymbol{y}^{(2)}=(b_{11},2b_{12},\cdots,2b_{1p},b_{22},\cdots,b_{pp})'$ 组成, 其中

$$\boldsymbol{B}=\sum_{\alpha=1}^N\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'\left(=\sum_{\alpha=1}^N\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'\right). \tag{17}$$

向量 $\boldsymbol{\omega}=(\boldsymbol{\omega}^{(1)\prime},\boldsymbol{\omega}^{(2)\prime})'$ 由 $\boldsymbol{\omega}^{(1)}=\sqrt{N}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 和 $\boldsymbol{\omega}^{(2)}=-\frac{1}{2}(\sigma^{11},\sigma^{12},\cdots,\sigma^{1p},\sigma^{22},\cdots,\sigma^{pp})'$ 组成. 测度 $m(A)$ 是集合 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N$ 映射到集合 A 的勒贝格测度.

引理 5.6.1 设 $\boldsymbol{B}=\boldsymbol{A}+N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'$. 则

$$N\bar{\boldsymbol{x}}'\boldsymbol{A}^{-1}\bar{\boldsymbol{x}}=\frac{N\bar{\boldsymbol{x}}'\boldsymbol{B}^{-1}\bar{\boldsymbol{x}}}{1-N\bar{\boldsymbol{x}}'\boldsymbol{B}^{-1}\bar{\boldsymbol{x}}}. \tag{18}$$

引理的证明 如果我们在 5.2 节的 (10) 中设 $\boldsymbol{B}=\boldsymbol{A}+\sqrt{N}\bar{\boldsymbol{x}}\sqrt{N}\bar{\boldsymbol{x}}'$, 则由推论 A.3.1 得到

$$\begin{aligned}\frac{1}{1+T^2/(N-1)}=\lambda^{2/N}&=\frac{|\boldsymbol{B}-\sqrt{N}\bar{\boldsymbol{x}}\sqrt{N}\bar{\boldsymbol{x}}'|}{|\boldsymbol{B}|}\\&=1-N\bar{\boldsymbol{x}}'\boldsymbol{B}^{-1}\bar{\boldsymbol{x}}.\end{aligned} \tag{19}$$

■

因此, 一个 T^2 检验的接受区域是

$$A=\{\boldsymbol{z}_N,\boldsymbol{B}|\boldsymbol{z}_N'\boldsymbol{B}^{-1}\boldsymbol{z}_N\leqslant k,\boldsymbol{B}\text{是正定的}\}, \tag{20}$$

对适当的 k.

当 $\boldsymbol{B}$ 正定时, 函数 $\boldsymbol{z}_N'\boldsymbol{B}^{-1}\boldsymbol{z}_N$ 是关于 $(\boldsymbol{z},\boldsymbol{B})$ 的凸函数 (习题 5.17). 因此, 集合 $\boldsymbol{z}_N'\boldsymbol{B}^{-1}\boldsymbol{z}_N\leqslant k$ 是凸集. 这说明了集合 A 是凸的. 进而, A 的闭包也是凸的 (习题 5.18), 且 A 的边界的概率是 0.

现在考虑定理 5.6.5 的其他条件. 假设 A 与半空间

$$c<\boldsymbol{\omega}'\boldsymbol{y}=\boldsymbol{\nu}'\boldsymbol{z}_N-\frac{1}{2}\operatorname{tr}\boldsymbol{\Lambda}\boldsymbol{B} \tag{21}$$

不相交, 其中 $\boldsymbol{\Lambda}$ 是一个对称矩阵且 $\boldsymbol{B}$ 是半正定矩阵. 我们取 $\boldsymbol{\Lambda}_1=\boldsymbol{I}$. 我们想证明 $\boldsymbol{\omega}_1+\lambda\boldsymbol{\omega}\in\Omega-\Omega_0$, 即 $\boldsymbol{\nu}_1+\lambda\boldsymbol{\nu}\neq 0$ (这是显然的) 和 $\boldsymbol{\Lambda}_1+\lambda\boldsymbol{\Lambda}$ 当 $\lambda>0$ 时正定. 这是当 $\boldsymbol{\Lambda}$ 为半正定时的情况. 现在我们将证明, 一个半空间 (21) 与 A 不相交和 $\boldsymbol{\Lambda}$ 不是半正定的是矛盾的. 如果 $\boldsymbol{\Lambda}$ 不是半正定的, 它可以写成 (由附录的推论 A.4.1)

$$\boldsymbol{\Lambda}=\boldsymbol{D}\begin{bmatrix}\boldsymbol{I}&0&0\\0&-\boldsymbol{I}&0\\0&0&0\end{bmatrix}\boldsymbol{D}', \tag{22}$$

其中 $\boldsymbol{D}$ 是非奇异的. 如果 $\boldsymbol{\Lambda}$ 不是半正定的, $-\boldsymbol{I}$ 是非空的, 因为它的阶数是 $\boldsymbol{\Lambda}$ 的负特征根的个数. 设 $\boldsymbol{z}_N = (1/\gamma)\boldsymbol{z}_0$ 且

$$\boldsymbol{B} = (\boldsymbol{D}')^{-1}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \gamma\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\boldsymbol{D}^{-1}. \tag{23}$$

则

$$\boldsymbol{\omega}'\boldsymbol{y} = \frac{1}{\gamma}\boldsymbol{\nu}'\boldsymbol{z}_0 + \frac{1}{2}\mathrm{tr}\begin{bmatrix} -\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \gamma\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tag{24}$$

对于充分大的 γ, 它大于 c. 另一方面

$$\boldsymbol{z}_N'\boldsymbol{B}^{-1}\boldsymbol{z}_N = \frac{1}{\gamma^2}\boldsymbol{z}_0'\boldsymbol{D}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \gamma^{-1}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\boldsymbol{D}'\boldsymbol{z}_0, \tag{25}$$

对于充分大的 γ, 它小于 k. 这和 (20) 与 (21) 不相交是矛盾的. 因此, 定理 5.6.5 的条件满足, 原定理得证. ■

这个证明来自 Stein.

容许性的另一个证明方法是证明 T^2 检验是一个真贝叶斯方法. 假定任意一个随机向量 $\boldsymbol{X}$ 有密度 $f(\boldsymbol{x}|\boldsymbol{\omega})$, 对 $\boldsymbol{\omega} \in \Omega$. 考虑检验原假设 $H_0: \boldsymbol{\omega} \in \Omega_0$ 及其备择假设 $H_1: \boldsymbol{\omega} \in \Omega - \Omega_0$. 设 Π_0 是 Ω_0 上的一个先验有限测度, Π_1 是 Ω_1 上的一个先验有限测度. 则贝叶斯方法 (及 $0-1$ 损失函数) 是拒绝 H_0, 如果对某个 c $(0 \leqslant c \leqslant \infty)$, 有

$$\frac{\int f(\boldsymbol{x}|\boldsymbol{\omega})\Pi_1(\mathrm{d}\boldsymbol{\omega})}{\int f(\boldsymbol{x}|\boldsymbol{\omega})\Pi_0(\mathrm{d}\boldsymbol{\omega})} \geqslant c. \tag{26}$$

如果 (26) 的等号对所有 $\boldsymbol{\omega} \in \Omega_0$ 的一个零测集成立, 则贝叶斯方法是唯一的, 从而是容许的. 因为这个测度有限, 他们可以正规为概率测度. 对于 $H_0: \boldsymbol{\mu} = \boldsymbol{0}$ 的 T^2 检验, 习题 5.15 提出了一对测度. (这对测度不是唯一的.) 读者可以用这些测度证明 (26) 可简化为 (20) 的补充.

在这些不变检验中, 已知 T^2 检验是一致最优的, 即给定显著性水平, 在所有不变检验中, 对于 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的每一个值, 它是最大功效的. 那么对于一个给定的 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 的值, T^2 检验是否是所有检验中“最好”的. 这里“最好”可以是容许的极小化极大; “极小化极大”是指极大化关于功效参数的极小值. Giri, Kiefer, and Stein (1963) 证明了 $p=2$ 和 $N=3$ 的最简单情况下的这个性质. 对于一般情况的 p 和 N, 这个性质是由 Šalaevskiĭ (1968) 给出的. 他给出了 $p=2$ 时的证明 [Šalaevskiĭ (1971)], 但是没有给出 $p>2$ 时的证明.

Giri and Kiefer (1964) 证明了 T^2 检验是局部极小化极大的 (当 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \to 0$) 且是渐近 (对数地) 极小化极大的, 当 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \to \infty$.

5.7 椭球等高分布

5.7.1 椭球等高观测

当 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})] \tag{1}$$

的容量为 N 的样本时, 样本均值 $\bar{\boldsymbol{x}}$ 和协方差阵 $\boldsymbol{S}$ 是分布均值 $\boldsymbol{\mu}=\boldsymbol{\nu}$ 和协方差阵 $\boldsymbol{\Sigma}=(\mathrm{E}(R^2/p))\boldsymbol{\Lambda}$ 的无偏估计, 其中 $R^2=(\boldsymbol{X}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\nu})$ 有有限期望. T^2 统计量 $T^2=N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})$ 可以用来检验和构造 $\boldsymbol{\mu}$ 的置信区域, 当 $\boldsymbol{\Sigma}$ (或 $\boldsymbol{\Lambda}$) 未知时, T^2 的小样本分布一般很难得到. 不过, 当 $N\to\infty$ 时 T^2 的极限分布可从 $\sqrt{N}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma})$ 和 $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Sigma}$ 得到 (定理 3.6.2).

定理 5.7.1 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自 (1) 的一组样本. 假设 $\mathrm{E}(R^2)<\infty$. 则 $T^2\xrightarrow{d}\chi^2_p$.

证明 定理 3.6.2 意味着 $N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})\xrightarrow{d}\chi^2_p$ 且 $N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})-T^2\xrightarrow{p}0$. ■

定理 5.7.1 说明, 对于椭球等高分布, 5.3 节的方法可以基于一组渐近基. 例如, 为了检验假设 $\boldsymbol{\mu}=\boldsymbol{\mu}_0$, 拒绝原假设, 如果

$$N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}_0)\geqslant\chi^2_p(\alpha), \tag{2}$$

其中 $\chi^2_p(\alpha)$ 是自由度为 p 的 χ^2 分布的 α 分位数, 当原假设为真且 $N\to\infty$ 时, (2) 的极限概率是 α. 类似地, 置信区域 $N(\bar{\boldsymbol{x}}-\boldsymbol{m})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{m})\leqslant\chi^2_p(\alpha)$ 有极限置信度 $1-\alpha$.

5.7.2 椭球等高矩阵分布

设 $\boldsymbol{X}(N\times p)$ 的密度

$$|\boldsymbol{C}|^{-N}g\left[\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')(\boldsymbol{C}')^{-1}\right] \tag{3}$$

基于左球面密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$. 这里 $\boldsymbol{Y}$ 有表示 $\boldsymbol{Y}\overset{d}{=}\boldsymbol{U}\boldsymbol{R}'$, 其中 $\boldsymbol{U}$ $(N\times p)$ 服从 O $(N\times p)$ 上的均匀分布, $\boldsymbol{R}$ 是下三角矩阵, 且 $\boldsymbol{U}$ 和 $\boldsymbol{R}$ 独立. 则 $\boldsymbol{X}\overset{d}{=}\boldsymbol{\varepsilon}_N\boldsymbol{\nu}'+\boldsymbol{U}\boldsymbol{R}'\boldsymbol{C}'$. 检验假设 $\boldsymbol{\nu}=\boldsymbol{0}$ 的 T^2 准则是 $N\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}$, 它关于变换 $\boldsymbol{x}\to\boldsymbol{X}\boldsymbol{G}$ 不变. 由推论 4.5.5, 我们得到下面的定理.

定理 5.7.2 假设 $\boldsymbol{X}$ 有密度 (3), 取 $\boldsymbol{\nu}=\boldsymbol{0}$ 和 $T^2=N\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}$. 则 $[T^2/(N-1)][(N-p)/p]$ 与 $F_{p,N-p}=(\chi^2_p/p)/[\chi^2_{N-p}/(N-p)]$ 同分布.

因此, 在给定显著性水平和置信度时, 这些假设检验和置信区域的构造对左球面分布是有效的.

$H:\boldsymbol{\nu}=\boldsymbol{0}$ 的 T^2 准则是

$$T^2 = N\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}} \stackrel{d}{=} N\bar{\boldsymbol{u}}'\boldsymbol{S}_{\boldsymbol{u}}^{-1}\bar{\boldsymbol{u}}, \tag{4}$$

这是因为 $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{U}\boldsymbol{R}'\boldsymbol{C}'$,

$$\bar{\boldsymbol{x}}' = \frac{1}{N}\boldsymbol{\varepsilon}_N'\boldsymbol{X} \stackrel{d}{=} \left(\frac{1}{N}\boldsymbol{\varepsilon}_N'\boldsymbol{U}\right)\boldsymbol{R}'\boldsymbol{C}' = \bar{\boldsymbol{u}}'(\boldsymbol{C}\boldsymbol{R})', \tag{5}$$

$$\begin{aligned}\boldsymbol{S} &= \frac{1}{N-1}(\boldsymbol{X}'\boldsymbol{X} - N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}') = \frac{1}{N-1}[\boldsymbol{C}\boldsymbol{R}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{R}\boldsymbol{C}' - \boldsymbol{C}\boldsymbol{R}\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}'(\boldsymbol{C}'\boldsymbol{R})'] \\ &= \boldsymbol{C}\boldsymbol{R}\boldsymbol{S}_{\boldsymbol{u}}(\boldsymbol{C}\boldsymbol{R})'.\end{aligned} \tag{6}$$

5.7.3 线性组合

Läuter, Glimm, and Kropf (1996a, 1996b, 1996c) 发现, 当 $\boldsymbol{\nu} = \boldsymbol{0}$ 时, 可以用 $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{C}\boldsymbol{R}\boldsymbol{R}'\boldsymbol{C}'$ 决定一个 $p \times q$ 的矩阵 $\boldsymbol{D}$ 且基于变换 $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{D}$ 的一个 T 检验. 特别地, 定义

$$\bar{\boldsymbol{z}}' = \frac{1}{N}\boldsymbol{\varepsilon}_N'\boldsymbol{Z} = \bar{\boldsymbol{x}}'\boldsymbol{D}, \tag{7}$$

$$\boldsymbol{S}_{\boldsymbol{Z}} = \frac{1}{N-1}(\boldsymbol{Z}'\boldsymbol{Z} - N\bar{\boldsymbol{z}}\bar{\boldsymbol{z}}') = \boldsymbol{D}'\boldsymbol{S}\boldsymbol{D}, \tag{8}$$

$$T_{\boldsymbol{D}}^2 = N\bar{\boldsymbol{z}}'S_2^{-1}\bar{\boldsymbol{z}}. \tag{9}$$

因为 $\boldsymbol{Q}_N\boldsymbol{Z} \stackrel{d}{=} \boldsymbol{Q}_N\boldsymbol{U}\boldsymbol{R}'\boldsymbol{C}' \stackrel{d}{=} \boldsymbol{U}\boldsymbol{R}'\boldsymbol{C}' = \boldsymbol{Z}$, 则矩阵 $\boldsymbol{Z}$ 基于左球面 $\boldsymbol{Y}\boldsymbol{D}$ 并因此有表示 $\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{R}^{*\prime}$, 其中 $\boldsymbol{V}$ $(N \times q)$ 服从 O $(N \times p)$ 上的均匀分布, 且与 $\boldsymbol{R}^{*\prime}$ (上三角矩阵) 独立, $\boldsymbol{R}^{*\prime}$ 的分布由 $\boldsymbol{R}^*\boldsymbol{R}^{*\prime} = \boldsymbol{Z}'\boldsymbol{Z}$ 得到. $T^2/(N-1)$ 的分布是 $F_{q,N-q}q/(N-q)$.

矩阵 $\boldsymbol{D}$ 也可以包含除 $\boldsymbol{X}'\boldsymbol{X}$ 之外的先验信息. 如果 p 比较大, q 可以比较小, 基于 $T_{\boldsymbol{D}}^2$ 的检验的功效可以比基于 T^2 的检验的功效更大.

Läuter, Glimm, and Kropf 给出一些选择 $\boldsymbol{D}$ 的例子. 其中一个是选择 $\boldsymbol{D}$ $(p \times 1)$ 为 $[\mathrm{Diag}\ (\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}]\boldsymbol{\epsilon}_p$, 其中 Diag $\boldsymbol{A}$ 是一个对角矩阵, 其第 i 个对角元是 a_{ii}. 统计量 $T_{\boldsymbol{D}}^2$ 称为标准和统计量.

习　题

5.1 (5.2 节) 设 $\boldsymbol{x}_\alpha$ 服从分布 $N(\boldsymbol{\mu} + \boldsymbol{\beta}(z_\alpha - \bar{z}), \boldsymbol{\Sigma})$, $\alpha = 1, \cdots, N$, 其中 $\bar{z} = (1/N)\sum z_\alpha$. 设 $\boldsymbol{b} = [1/\sum(z_\alpha - \bar{z})^2]\sum \boldsymbol{x}_\alpha(z_\alpha - \bar{z})$, $(N-2)\boldsymbol{S} = \sum[\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}} - \boldsymbol{b}(z_\alpha - \bar{z})][\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}} - \boldsymbol{b}(z_\alpha - \bar{z})]'$, 且 $T^2 = \sum(z_\alpha - \bar{z})^2\boldsymbol{b}'\boldsymbol{S}^{-1}\boldsymbol{b}$. 证明 T^2 服从自由度为 $N-2$ 的 T^2 分布. [提示: 见习题 3.13.]

5.2 (5.2.2 节) 证明 $T^2/(N-1)$ 可以写成 $R^2/(1-R^2)$, 其对应关系如表 5.1 所示.

5.3 (5.2.2 节) 设

$$\frac{R^2}{1-R^2} = \frac{\sum u_\alpha \boldsymbol{x}_\alpha'(\sum \boldsymbol{x}_\alpha\boldsymbol{x}_\alpha')^{-1}\sum u_\alpha\boldsymbol{x}_\alpha}{\sum u_\alpha^2 - \sum u_\alpha\boldsymbol{x}_\alpha'(\sum \boldsymbol{x}_\alpha\boldsymbol{x}_\alpha')^{-1}\sum u_\alpha\boldsymbol{x}_\alpha},$$

其中 $u_1, \cdots, u_N$ 是 N 个数且 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 独立服从分布 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$. 证明 $R^2/(1-R^2)$ 的分布独立于 $u_1, \cdots, u_N$. [提示: 存在一个 $N \times N$ 正交矩阵 $\boldsymbol{C}$ 将 $(u_1, \cdots, u_N)$ 变换为等比于 $(1/\sqrt{N}, \cdots, 1/\sqrt{N})$ 的向量.]

表 5.1

5.2 节	4.4 节
$x_{0\alpha}=1/\sqrt{N}$	$z_{1\alpha}$
$\boldsymbol{x}_\alpha$	$z_\alpha^{(2)}$
$\sqrt{N}\bar{\boldsymbol{x}}$	$\boldsymbol{a}_{(1)}=\sum z_{1\alpha}\boldsymbol{z}_\alpha^{(2)}$
$\boldsymbol{B}=\sum \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha'$	$\boldsymbol{A}_{22}=\sum z_\alpha^{(2)} z_\alpha^{(2)\prime}$
$1=\sum x_{0\alpha}^2$	$a_{11}=\sum z_{1\alpha}^2$
$\frac{T^2}{N-1}$	$\frac{R^2}{1-R^2}$
p	$p-1$
N	n

5.4 (5.2.2 节) 利用习题 5.2 和习题 5.3 证明 $[T^2/(N-1)][(N-p)/p]$ 服从 $F_{p,N-p}$ 分布 (在原假设下). [注意: 这是对应于 Hotelling 几何证明 (1931) 的分析.]

5.5 (5.2.2 节) 设 $T^2=N\bar{\boldsymbol{x}}'\boldsymbol{S}^{-1}\bar{\boldsymbol{x}}$, 其中 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的容量为 N 的样本的均值向量和协方差阵. 证明当用 $\boldsymbol{\lambda}=(\tau,0,\cdots,0)'$ 替换 $\boldsymbol{\mu}$ (其中 $\tau^2=\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$) 和用 $\boldsymbol{I}$ 替换 $\boldsymbol{\Sigma}$ 时, T^2 的分布相同.

5.6 (5.2.2 节) 设 $u=[T^2/(N-1)]/[1+T^2/(N-1)]$. 证明 $u=\boldsymbol{\gamma}\boldsymbol{V}'(\boldsymbol{V}\boldsymbol{V}')^{-1}\boldsymbol{V}\boldsymbol{\gamma}'$, 其中 $\boldsymbol{\gamma}=(1/\sqrt{N},\cdots,1/\sqrt{N})$ 且

$$\boldsymbol{V}=\begin{pmatrix}\boldsymbol{v}_1\\ \vdots\\ \boldsymbol{v}_p\end{pmatrix}=\begin{pmatrix}x_{11} & \cdots & x_{1N}\\ \vdots & & \vdots\\ x_{p1} & \cdots & x_{pN}\end{pmatrix}.$$

5.7 (5.2.2 节) 设

$$\begin{aligned}
\boldsymbol{v}_1^* &= \boldsymbol{v}_1\\
\boldsymbol{v}_i^* &= \boldsymbol{v}_i-\frac{\boldsymbol{v}_i\boldsymbol{v}_1'}{\boldsymbol{v}_1\boldsymbol{v}_1'}\boldsymbol{v}_1=\boldsymbol{v}_1\left(\boldsymbol{I}-\frac{1}{\boldsymbol{v}_1\boldsymbol{v}_1'}\boldsymbol{v}_1'\boldsymbol{v}_1\right), \quad i\neq 1,\\
\boldsymbol{\gamma}^* &= \boldsymbol{\gamma}-\frac{\boldsymbol{\gamma}\boldsymbol{v}_1'}{\boldsymbol{v}_1\boldsymbol{v}_1'}\boldsymbol{v}_1,\\
\boldsymbol{V}^* &= \begin{pmatrix}\boldsymbol{v}_1^*\\ \vdots\\ \boldsymbol{v}_p^*\end{pmatrix}.
\end{aligned}$$

证明 $U=s+(1-s)w$, 其中

$$s=\frac{(\boldsymbol{\gamma}\boldsymbol{v}_1^{*\prime})^2}{\boldsymbol{v}_1^*\boldsymbol{v}_1^{*\prime}}=\frac{(\boldsymbol{\gamma}\boldsymbol{v}_1')^2}{\boldsymbol{v}_1\boldsymbol{v}_1'},$$

$$w=\frac{1}{\boldsymbol{\gamma}^*\boldsymbol{\gamma}^{*\prime}}\boldsymbol{\gamma}^*\begin{pmatrix}\boldsymbol{v}_2^*\\ \vdots\\ \boldsymbol{v}_p^*\end{pmatrix}'\begin{pmatrix}\boldsymbol{v}_2^*\boldsymbol{v}_2^{*\prime} & \cdots & \boldsymbol{v}_2^*\boldsymbol{v}_p^{*\prime}\\ \vdots & & \vdots\\ \boldsymbol{v}_p^*\boldsymbol{v}_2^{*\prime} & \cdots & \boldsymbol{v}_p^*\boldsymbol{v}_p^{*\prime}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{v}_2^*\\ \vdots\\ \boldsymbol{v}_p^*\end{pmatrix}\boldsymbol{\gamma}^{*\prime}.$$

提示: $\boldsymbol{E}\boldsymbol{V}=\boldsymbol{V}^*$, 其中

$$E = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\frac{v_2 v_1'}{v_1 v_1'} & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ -\frac{v_p v_1'}{v_1 v_1'} & 0 & \cdots & 1 \end{pmatrix}.$$

5.8 (5.2.2 节) 对于 $(N-1)$ 维空间上一个向量和 $p-1$ 个向量的不去均值的多重相关系数, w 的分布和该相关系数的分布相同即 w 有密度

$$\frac{\Gamma[\frac{1}{2}(N-1)]}{\Gamma[\frac{1}{2}(N-p)]\Gamma[\frac{1}{2}(p-1)]} w^{\frac{1}{2}(p-1)-1}(1-w)^{\frac{1}{2}(N-p)-1}.$$

[提示: 习题 5.7 的变换是 $v_2, \cdots, v_p, \gamma$ 在正交于 v_1 的 $(N-1)$ 维空间上的投影.]

5.9 (5.2.2 节) 证明 $r = s/(1-s)$ 乘以 $(N-1)/1$ 服从自由度为 1 和 $N-1$ 且非中心参数为 $N\tau^2$ 的非中心 F 分布.

5.10 (5.2.2 节) 由习题 5.5~5.9, 证明推论 5.2.1.

5.11 (5.3 节) 利用 3.2 节的数据, 求假设 "没有一种药物有催眠效应" 的显著性水平为 0.01 的检验.

5.12 (5.3 节) 用 3.2 节的数据, 给出 $\boldsymbol{\mu}$ 的置信度为 0.95 的置信区域.

5.13 (5.3 节) 证明 5.3.6 节的论述: T^2 统计量独立于 $\boldsymbol{C}$ 的选择.

5.14 (5.5 节) 用习题 4.41 的数据求假设 "长子的平均头长和头宽等于次子的平均头长和头宽" 的显著性水平为 0.01 的检验.

5.15 (5.6.2 节) *贝叶斯方法的 T^2 检验* [Kiefer and Schwartz (1965)]. 设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 独立服从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 分布. 定义 Π_0 为 $[\boldsymbol{\mu}, \boldsymbol{\Sigma}] = [\boldsymbol{0}, (\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-1}]$, 其中 $\boldsymbol{\eta}$ 的密度等比于 $|\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}'|^{-\frac{1}{2}N}$, 定义 Π_1 为 $[\boldsymbol{\mu}, \boldsymbol{\Sigma}] = [(\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-1}\boldsymbol{\eta}, (\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-1}]$, 其中 $\boldsymbol{\eta}$ 的密度等比于

$$|\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}'|^{-\frac{1}{2}N} \exp\left[\frac{1}{2}N\boldsymbol{\eta}'(\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-1}\boldsymbol{\eta}\right].$$

(a) 对 $N > p$, 通过证明 $\boldsymbol{\eta}'(\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-1}\boldsymbol{\eta} \leqslant 1$ 和 $|\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}'|^{-\frac{1}{2}N} = (\boldsymbol{I}+\boldsymbol{\eta}\boldsymbol{\eta}')^{-\frac{1}{2}N}$ 的积分有限来证明这些测度是有限的.

(b) 证明不等式 (26) 等价于 $N\bar{\boldsymbol{x}}'(\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha')^{-1}\bar{\boldsymbol{x}} \geqslant k$. 因此, T^2 检验是贝叶斯的因而是容许的.

5.16 (5.6.2 节) 设 $g(t) = f[t\boldsymbol{y}_1 + (1-t)\boldsymbol{y}_2]$, 其中 $f(\boldsymbol{y})$ 是向量 $\boldsymbol{y}$ 的一个实值函数. 证明如果 $g(t)$ 是凸的, 则 $f(\boldsymbol{y})$ 也是凸的.

5.17 (5.6.2 节) 证明 $\boldsymbol{z}'\boldsymbol{B}^{-1}\boldsymbol{z}$ 是 $(\boldsymbol{z}, \boldsymbol{B})$ 的一个凸函数, 其中 $\boldsymbol{B}$ 是正定矩阵. [提示: 利用习题 5.16.]

5.18 (5.6.2 节) 证明如果集合 $\boldsymbol{A}$ 是凸的, 则 $\boldsymbol{A}$ 的闭包也是凸的.

5.19 (5.3 节) 设 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 基于来自总体 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 N 个观测. 证明 $\boldsymbol{x} - \bar{\boldsymbol{x}}$ 服从分布

$$N[\boldsymbol{0}, (1+1/N)\boldsymbol{\Sigma}].$$

证明 $[N/(N+1)](\boldsymbol{x}-\bar{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})$ 服从自由度为 $N-1$ 的 T^2 分布. 说明这个统计量怎样基于 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 给出 $\boldsymbol{x}$ 的一个预测区域. (即使得下一个观测落在其上的概率为指定置信度的区域).

5.20 (5.3 节) 设 $\boldsymbol{x}_\alpha^{(i)}$ 是来自 $N(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma}_i)$ $(\alpha=1,\cdots,N_i, i=1,2)$ 的观测. 求假设检验 $\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}$ 的似然比准则.

5.21 (5.4 节) 通过验证

$$\frac{1}{1-\rho^2}\left(\frac{\mu_1^2}{\sigma_1^2}-2\rho\frac{\mu_1\mu_2}{\sigma_1\sigma_2}+\frac{\mu_2^2}{\sigma_2^2}\right)=\frac{\mu_1^2}{\sigma_1^2}+\frac{(\mu_2-\rho\sigma_2\mu_1/\sigma_1)^2}{(1-\rho^2)\sigma_2^2}$$

来证明当 $\boldsymbol{\mu}'=(\mu_1,\mu_2)$ 时的 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ 大于当 $\boldsymbol{\mu}=\mu_1$ 时的 $\boldsymbol{\mu}'\boldsymbol{\Sigma}'\boldsymbol{\mu}$. 比较检验 $\mu_1=0$ 的功效和检验 $\mu_1=0, \mu_2=0$ 的功效.

5.22 (5.3 节)

(a) 用 5.3.4 节的数据检验假设 $\mu_1^{(1)}=\mu_1^{(2)}$.

(b) 检验假设 $\mu_1^{(1)}=\mu_1^{(2)}, \mu_2^{(1)}=\mu_2^{(2)}$.

5.23 (5.4 节) 设

$$\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}$$

证明 $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\geqslant\boldsymbol{\mu}^{(1)'}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}^{(1)}$. 给出不等式严格成立的一个条件. [提示: 这是类比习题 5.21 的向量.]

5.24 设 $\boldsymbol{X}^{(i)'}=(\boldsymbol{Y}^{(i)'}, \boldsymbol{Z}^{(i)'})$ $(i=1,2$, 其中 $\boldsymbol{Y}^{(i)}$ 有 p 个分量, $\boldsymbol{Z}^{(i)}$ 有 q 个分量) 服从分布 $N(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma})$, 其中

$$\boldsymbol{\mu}^{(i)}=\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{y}}^{(i)}\\ \boldsymbol{\mu}_{\boldsymbol{z}}^{(i)}\end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{yy}} & \boldsymbol{\Sigma}_{\boldsymbol{yz}}\\ \boldsymbol{\Sigma}_{\boldsymbol{zy}} & \boldsymbol{\Sigma}_{\boldsymbol{zz}}\end{pmatrix},\quad i=1,2.$$

基于 $\boldsymbol{X}^{(i)}$ 上容量为 N_i 的样本, $i=1,2$, 求检验给定 $\boldsymbol{\mu}_{\boldsymbol{z}}^{(1)}=\boldsymbol{\mu}_{\boldsymbol{z}}^{(2)}$ 时 $\boldsymbol{\mu}_{\boldsymbol{y}}^{(1)}=\boldsymbol{\mu}_{\boldsymbol{y}}^{(2)}$ 的似然比准则. [提示: 用 $\boldsymbol{Y}^{(i)}$ 的边缘密度和给定 $\boldsymbol{Y}^{(i)}$ 时 $\boldsymbol{Z}^{(i)}$ 的条件密度来表示这个似然.]

5.25 求前面问题中的准则在原假设下的分布.

5.26 (5.5 节) 假设 $\boldsymbol{x}_\alpha^{(g)}$ 是来自 $N(\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g)$ $(\alpha=1,\cdots,N_g, g=1,\cdots,q)$ 的一个观测,

(a) 证明假设 $\boldsymbol{\mu}^{(1)}=\cdots=\boldsymbol{\mu}^{(q)}$ 等价于 $\mathrm{E}(\boldsymbol{y}_\alpha^{(i)})=\mathbf{0}$, $i=1,\cdots,q-1$, 其中

$$\boldsymbol{y}_\alpha^{(i)}=a_1^{(i)}\boldsymbol{x}_\alpha^{(1)}+\sum_{g=2}^{q}a_g^{(i)}\left(\frac{N_1}{N_g}\right)^{\frac{1}{2}}\left[\boldsymbol{x}_\alpha^{(g)}-\frac{1}{N_1}\sum_{\beta=1}^{N_1}\boldsymbol{x}_\beta^{(g)}+\frac{1}{(N_1N_g)^{\frac{1}{2}}}\sum_{\beta=1}^{N_g}\boldsymbol{x}_\beta^{(g)}\right],$$
$$\alpha=1,\cdots,N_1, i=1,\cdots,q-1;$$

$N_1\leqslant N_g, g=2,\cdots,q$; $(a_1^{(i)},\cdots,a_q^{(i)}), i=1,\cdots,q-1$, 是线性独立的.

(b) 说明怎样用 $(\bar{\boldsymbol{y}}^{(1)'},\cdots,\bar{\boldsymbol{y}}^{(q-1)'})'$ 构造假设的 T^2 检验, 它产生一个自由度为 $(q-1)p$ 和 $N-(q-1)p$ 的 F 统计量 [Anderson (1963b)].

5.27 (5.2 节) 证明 (25) 是 $V=\chi_a^2/(\chi_a^2+\chi_b^2)$ 的密度. [提示: 在 $U=\chi_a^2$ 和 $W=\chi_b^2$ 的联合密度里做变换 $u=vw(1-v)^{-1}, w=w$, 并且对 w 积分.]

第6章　观察值的分类

6.1　分 类 问 题

当研究人员拥有一些个体的观测数据, 并希望利用这些数据确定该个体的所属类别时, 分类问题[1]就产生了. 因为必须借助于观测数据才能将个体归类, 我们一般假设个体来自于有限个可能的类别或总体, 而且每个总体由各自的概率分布决定. 把研究个体看作是来自总体的随机观测, 分类问题就变为: 在给定观测值的条件下, 个体来自哪个总体?

分类问题可以看作是“统计决策函数”问题. 我们对观测值的分布作出许多假设, 但只能接受其中之一. 只有两个备选总体的情况就是最基本的假设检验问题, 即接受其中一个假设而拒绝另一个.

在有些情况, 各总体的概率分布被当作是完全已知的, 因而类别是预先确定的; 而在其他情况下, 分布的形式也许是已知但其中的参数必须由样本估计而得.

下面是一个分类问题的例子. 申请进入大学的学生都要接受一系列的测试, 测试结果组成的集合是一个得分向量 $\boldsymbol{x}$. 每个申请者都有可能顺利完成或不能顺利完成大学学业. 现在的问题就是如何根据入学考试成绩对申请者进行分类.

本章介绍分类理论的一般性结论, 并将其应用在总体服从正态分布的情况. 6.2 节在决策理论的基础上给出了两总体分类问题的定义, 6.3 节介绍了贝叶斯判别过程和可容许解. 6.4 节将分类理论应用于两个均值不同的已知正态分布总体, 得到了总体的线性判别函数 (discriminant function). 当总体参数未知时, 使用其估计值 (6.5 节). 另一种方法是极大似然方法. 6.6 节利用分布函数的渐近展开, 分别估计出两种分类方法的误判概率. 接着这些结果被应用于多总体情况. 最后, 6.10 节研究了参数已知时协方差阵不同的两总体的线性判别问题.

6.2　精确分类的标准

6.2.1　预备知识

我们希望分类方法能够极小化误判概率, 或者说极小化因误判而导致的平均损失. 下面将详细地叙述这个问题. 为叙述方便, 现在假设只有两个类别, 一般的情况

① 也叫判别问题. —— 译者注

稍后讨论. 本节内容是 3.4 节的两决策问题的详细推广.

假设个体是来自总体 π_1 或总体 π_2 的观测. 测量向量 $x' = (x_1, \cdots, x_p)$ 被用来判断个体所属类别. 我们建立一个准则, 若个体的观测值 $x_1, \cdots, x_p$ 属于某集合, 就认为这个个体来自 π_1, 否则来自 π_2.

把每个观测都看作 p 维空间的一个点, 将 p 维空间分成两个区域, R_1, R_2. 如果观测落入区域 R_1, 就认为它来自总体 π_1; 如果观测落入区域 R_2, 就认为它来自总体 π_2.

在下面给定的分类方法中, 统计学家可能犯两种错误. 第一种错误是把来自 π_1 的个体误判为来自 π_2, 第二种错误是把来自 π_2 的个体误判为来自 π_1. 我们需要知道这两种错误哪一个更加严重. 假设第一种错误和第二种错误造成的损失分别为 $C(2|1)(>0)$ 和 $C(1|2)(>0)$. 这些损失可以用任何单位度量. 在后面我们会看到, 真正重要的是两种损失的比值. 统计学家可能不知道这两种错误造成的精确损失, 但是通常会对它们有一个粗略的了解.

表 6.1 中给出了正确分类和错误分类的损失. 很明显, 一个好的分类方法应该在某种意义上极小化误判损失.

表 6.1

		统计学家的决策	
		π_1	π_2
总体	π_1	0	$C(2\|1)$
	π_2	$C(1\|2)$	0

6.2.2 两总体的两种情况

我们将以两种方式定义“极小损失”. 第一种情况假设两个总体的先验概率已知, 观测来自总体 π_1 和 π_2 的概率分别为 q_1 和 $q_2(q_1+q_2=1)$, 总体 π_1 的概率性质由分布函数决定. 为方便记, 我们总是假设分布函数具有密度函数, 因为离散的情况类似. 设总体 π_1 和 π_2 的密度函数分别为 $p_1(x)$ 和 $p_2(x)$. 如果当观测值落入区域 R_1 时我们就认为该观测来自于总体 π_1, 那么把来自 π_1 的观测正确分类的概率为

$$P(1|1,R) = \int_{R_1} p_1(x)\mathrm{d}\boldsymbol{x}, \tag{1}$$

其中 $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1 \cdots \mathrm{d}x_p$, 误判的概率为

$$P(2|1,R) = \int_{R_2} p_1(x)\mathrm{d}\boldsymbol{x}. \tag{2}$$

同理, 把来自总体 π_2 的观测正确分类的概率为

$$P(2|2,R) = \int_{R_2} p_2(x)\mathrm{d}\boldsymbol{x}, \tag{3}$$

误判的概率为

$$P(1|2,R) = \int_{R_1} p_2(x)\mathrm{d}\boldsymbol{x}. \tag{4}$$

由于任给一个观测它来自总体 π_1 的概率为 q_1, 所以我们把一个来自 π_1 的观测正确分类的概率为 $q_1P(1|1,R)$, 这就是表 6.1 左上角所对应情况的概率; 我们把这个观测误判的概率为 $q_1P(2|1,R)$. 同理可得, 表 6.1 左下角和右下角所对应的概率分别为 $q_2P(1|2,R)$, $q_2P(2|2,R)$.

判断错误的平均损失或期望损失是什么呢? 它就是各种误判概率与其损失乘积的和, 即

$$C(2|1)P(2|1,R)q_1 + C(1|2)P(1|2,R)q_2. \tag{5}$$

这就是我们要极小化的平均损失. 也就是说, 我们希望把整个空间分成 R_1 和 R_2 两部分, 使得上面的期望损失达到极小. 在给定 q_1 和 q_2 的前提下极小化 (5) 的判别方法 (或分类方法) 称为*贝叶斯判别 (或分类)*.

在申请入学的例子中, 两种误判情况分别是招收了一个不能顺利完成学业的学生和拒绝了一个有潜质的好学生.

还有一些情况, 我们不知道先验概率. 这时观测来自 π_1 的期望损失是

$$C(2|1)P(2|1,R) = r(1,R); \tag{6}$$

观测来自 π_2 的期望损失是

$$C(1|2)P(1|2,R) = r(2,R). \tag{7}$$

我们不知道观测是来自于 π_1 还是 π_2, 也不知道这两种情况的概率.

对于判别程式 (或方法) R 和 R^*, 如果 $r(1,R) \leqslant r(1,R^*)$, $r(2,R) \leqslant r(2,R^*)$, 则认为程式 R 优于程式 R^*. 如果上面两个不等式至少有一个取严格不等号, 我们就说程式 R 严格优于程式 R^*. 一般情况下, 不存在某一判别程式严格优于或优于任何其他的程式. 若不存在其他的程式优于程式 R, 我们就称程式 R 是*容许的*, 我们只对容许判别程式组成的集合感兴趣. 下面我们将会看到, 在一些限制条件下, 这个集合与贝叶斯判别所组成的集合是相同的. 设 R 是由一些判别程式组成的集合, 那么: (1) 如果对于 R 以外的任何一个程式, 都存在 R 内的某一程式严格优于这个程式, 则称 R 为*完全的或完全类*; (2) 如果对于 R 以外的任何一个程式, 都存在 R 内的某一程式优于这个程式, 则称 R 为*本质完全的或本质完全类*; (3) 若 R 是完全类, 且对于任意的 $R_1 \subset R, R_1$ 都不是完全的, 则称 R 为*最小完全类*. 同理可以定义*最小本质完全类*. 下面我们将会看到, 在某些条件下, 可容许类就是最小完全类. 为方便叙述, 仅相差零测集的判别程式被认为是一样的. 不特殊说明的话, 下节的结论都是对非零测集而言的.

一个判别称为*极小化极大判别*, 若它的极大期望损失 $r(i,R)$ 是极小的. 一般情况下, 通过极小化极大原则可以找到唯一的判别. 从保守的观点来看, 这种方法选出的判别应是最好的. 关于这部分的详细讨论请参阅 Wald(1950), Blackell and Girshick (1954), Ferguson (1967), DeGroot (1970) 和 Berger(1980b).

6.3　概率分布已知的两总体的判别

6.3.1　先验分布已知的情况

我们现在讨论如何选取区域 R_1, R_2, 以便 6.2 节中的 (5) 式达到极小. 因为先验概率是已知的, 所以我们可以定义总体和变量观测集的联合概率. 若观测来自 π_1, 则每个变量取值不大于 y 的相应分量的概率为

$$\int_{-\infty}^{y_p} \cdots \int_{-\infty}^{y_1} q_1 p_1(x) \mathrm{d}x_1 \cdots \mathrm{d}x_p. \tag{1}$$

我们同样可以定义在给定观测变量值的条件下观测来自某一总体的概率. 例如, 给定观测 $\boldsymbol{x}$, 其来自总体 π_1 的条件概率为

$$\frac{q_1 p_1(\boldsymbol{x})}{q_1 p_1(\boldsymbol{x}) + q_2 p_2(\boldsymbol{x})}. \tag{2}$$

若 $C(1|2) = C(2|1) = 1$, 则期望损失为

$$q_1 \int_{R_2} p_1(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + q_2 \int_{R_1} p_2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{3}$$

显然, 上式也是误判概率, 因此我们希望极小化误判概率.

给定观测点 $\boldsymbol{x}$, 我们通过把它分派给条件概率较高的总体来极小化误判概率. 如果

$$\frac{q_1 p_1(\boldsymbol{x})}{q_1 p_1(\boldsymbol{x}) + q_2 p_2(\boldsymbol{x})} \geqslant \frac{q_2 p_2(\boldsymbol{x})}{q_1 p_1(\boldsymbol{x}) + q_2 p_2(\boldsymbol{x})}, \tag{4}$$

我们认为观测来自总体 π_1, 否则我们认为观测来自总体 π_2. 因为误判概率在每个点上极小化, 因此在整个空间上达到极小, 从而判定准则变为

$$\begin{aligned} R_1 &: q_1 p_1(\boldsymbol{x}) \geqslant q_2 p_2(\boldsymbol{x}), \\ R_2 &: q_1 p_1(\boldsymbol{x}) < q_2 p_2(\boldsymbol{x}). \end{aligned} \tag{5}$$

如果 $q_1 p_1(\boldsymbol{x}) = q_2 p_2(\boldsymbol{x})$, 那么我们既可以认为这个观测来自 π_1 也可以认为它来自 π_2, 这里我们不妨把这个点归入 R_1. 对给定的 $\boldsymbol{x}$, 如果 $q_1 p_1(\boldsymbol{x}) + q_2 p_2(\boldsymbol{x}) = 0$, 则我们把这个点分入两个区域的任何一个都可以.

下面我们严格证明 (5) 就是最佳判别. 任给判别 $R^* = (R_1^*, R_2^*)$, 误判概率为

$$\begin{aligned} &q_1 \int_{R_2^*} p_1(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + q_2 \int_{R_1^*} p_2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \int_{R_2^*} [q_1 p_1(\boldsymbol{x}) - q_2 p_2(\boldsymbol{x})] \mathrm{d}\boldsymbol{x} + q_2 \int p_2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{6}$$

显然, 上面右边式子的第二项是常数; 如果 R_2^* 包含所有满足条件 $q_1 p_1(\boldsymbol{x}) - q_2 p_2(\boldsymbol{x}) < 0$ 的点并且不包含所有使得 $q_1 p_1(\boldsymbol{x}) - q_2 p_2(\boldsymbol{x}) > 0$ 的点, 则第一项可以达到极小. 如果

$$\Pr\left\{\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}=\frac{q_2}{q_1}\Big|\pi_i\right\}=0,\quad i=1,2, \tag{7}$$

则贝叶斯判别在零概率集以外的情况下是唯一的.

现在我们只需关心下面的数学问题：给定非负常数 q_1,q_2 及非负函数 $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$, 选择区域 R_1,R_2 使 (3) 式达到极小. 由上面的讨论知它的解就是 (5) 式. 如果我们想极小化 6.2 节中的 (5) 式, 即极小化

$$[C(2|1)q_1]\int_{R_2}p_1(\boldsymbol{x})\mathrm{d}\boldsymbol{x}+[C(1|2)q_2]\int_{R_1}p_2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{8}$$

由于 $C(2|1)q_1$ 和 $C(1|2)q_2$ 都是非负常量, 则我们可根据下式选择的 R_1 和 R_2:

$$\begin{aligned}R_1&:[C(2|1)q_1]p_1(\boldsymbol{x})\geqslant[C(1|2)q_2]p_2(\boldsymbol{x}),\\R_2&:[C(2|1)q_1]p_1(\boldsymbol{x})<[C(1|2)q_2]p_2(\boldsymbol{x}).\end{aligned} \tag{9}$$

(9) 式可以改写为

$$\begin{aligned}R_1&:\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\geqslant\frac{C(1|2)q_2}{C(2|1)q_1},\\R_2&:\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}<\frac{C(1|2)q_2}{C(2|1)q_1}.\end{aligned} \tag{10}$$

定理 6.3.1 如果观测来自总体 π_1 和 π_2 的先验概率分别为 q_1 和 q_2, 并设 $p_1(\boldsymbol{x})$ 和 $p_2(\boldsymbol{x})$ 分别为 π_1 和 π_2 的密度, 来自 π_1 的观测误判为 π_2 的损失为 $C(2|1)$, 来自 π_2 的观测误判为 π_1 的损失为 $C(1|2)$, 则由 (10) 式定义的分类区域 R_1 和 R_2 极小化期望损失. 如果

$$\Pr\left\{\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}=\frac{C(1|2)q_2}{C(2|1)q_1}\Big|\pi_i\right\}=0,\quad i=1,2, \tag{11}$$

则上面的判别程式在零概率集之外的情况下是唯一的.

6.3.2 先验分布未知的情况

在许多情况下, 统计学家无法确定两个总体的先验概率. 这时我们将寻找容许判别程式, 也就是最优判别程式. 首先我们来证明贝叶斯判别是容许的. 设 $R=(R_1,R_2)$ 是一个贝叶斯判别, 且观测来自两个总体的先验概率分别为 q_1,q_2. 是否存在过程 $R^*=(R_1^*,R_2^*)$ 使得 $P(1|2,R^*)\leqslant P(1|2,R)$ 和 $P(2|1,R^*)\leqslant P(2|1,R)$, 并且这两个不等式至少有一个严格成立? 因为 R 是贝叶斯判别, 所以

$$q_1P(2|1,R)+q_2P(1|2,R)\leqslant q_1P(2|1,R^*)+q_2P(1|2,R^*). \tag{12}$$

这个不等式可以写为

$$q_1[P(2|1,R)-P(2|1,R^*)]\leqslant q_2[P(1|2,R^*)-P(1|2,R)]. \tag{13}$$

假设 $0<q_1<1$. 如果 $P(1|2,R^*)\leqslant P(1|2,R)$, 则上式的右边小于零, 即 $P(2|1,R)\leqslant P(2|1,R^*)$, 此时 R^* 并不优于 R. 若 $P(2|1,R^*)\leqslant P(2|1,R)$, 同理可推出 $P(1|2,R)\leqslant$

$P(1|2,R^*)$, 此时 R^* 也不优于 R. 假设 $q_1=0$, 由 (13) 式可得 $0\leqslant P(1|2,R^*)-P(1|2,R)$. 又由于 R 是贝叶斯判别, 则 R_1 只包含满足条件 $p_2(\boldsymbol{x})=0$ 的点[1], 因此 $P(1|2,R)=0$. 如果 R^* 优于 R, 则 $P(1|2,R^*)=0$. 若 $\Pr\{p_2(\boldsymbol{x})=0|\pi_1\}=0$, 则 $P(2|1,R)=\Pr\{p_2(\boldsymbol{x})>0|\pi_1\}=1$. 若 $P(1|2,R^*)=0$, 则 R_1^* 只包含满足条件 $p_2(\boldsymbol{x})=0$ 的点. 从而 $P(2|1,R^*)=\Pr\{R_2^*|\pi_1\}=\Pr\{p_2(\boldsymbol{x})>0|\pi_1\}=1$, 此时 R^* 也不优于 R.

定理 6.3.2 *如果 $\Pr\{p_2(\boldsymbol{x})=0|\pi_1\}=0$ 和 $\Pr\{p_1(\boldsymbol{x})=0|\pi_2\}=0$ 成立, 则所有的贝叶斯判别都是容许的.*

下面我们证明上面定理的逆问题, 即每个容许判别程式都是贝叶斯判别. 假设[2]

$$\Pr\left\{\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}=k\middle|\pi_i\right\}=0,\quad i=1,2,0\leqslant k\leqslant\infty. \tag{14}$$

则对任意的 q_1, 贝叶斯判别是唯一的. 另外, 在给定 π_1 或 π_2 的条件下, $\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}$ 的分布函数是连续的.

设 R 是容许判别程式, 则存在 k 满足

$$P(2|1,R)=\Pr\left\{\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\leqslant k\middle|\pi_i\right\}=P(2|1,R^*), \tag{15}$$

其中 R^* 是对应于 $q_2/q_1=k$[即 $q_1=1/(1+k)$] 的贝叶斯判别. 因为 R 是容许的, 所以 $P(1|2,R)\leqslant P(1|2,R^*)$. 然而, 由定理 6.3.2 可知, R^* 也是容许的, 所以 $P(1|2,R)\geqslant P(1|2,R^*)$, 从而 $P(1|2,R)=P(1|2,R^*)$. 于是 R 也是贝叶斯判别, 由贝叶斯判别的唯一性可知程式 R 与程式 R^* 相同.

定理 6.3.3 *如果 (14) 式成立, 则每一个容许判别程式都是贝叶斯判别.*

定理 6.3.3 的证明说明贝叶斯判别组成的类是完全类. 任给这个类外面的一个判别程式 R, 我们都可以构造一个贝叶斯判别 R^*, 使得 $P(2|1,R)=P(2|1,R^*)$. 由于 R^* 是容许的, 则 $P(1|2,R)\geqslant P(1|2,R^*)$. 另外, 由于贝叶斯判别类与容许判别程式类相同, 所以贝叶斯判别类是最小完全类.

定理 6.3.4 *如果 (14) 式成立, 则贝叶斯判别类是最小完全类.*

最后, 我们来讨论极小化极大判别程式. 设 $P(i|j,q_1)=P(i|j,R)$, 其中 R 是对应于 q_1 的贝叶斯判别. $P(i|j,q_1)$ 是 q_1 的连续函数. 当 q_1 从 0 变化到 1 时, $P(2|1,q_1)$ 从 1 变化到 0; $P(1|2,q_1)$ 从 0 变化到 1. 因此存在 q_1 的一个值, 不妨设为 q_1^*, 满足 $P(2|1,q_1^*)=P(1|2,q_1^*)$. 这就是极小化极大解, 因为如果有另外一个判别程式 R^* 满足 $\max\{P(2|1,R^*),P(1|2,R^*)\}\leqslant P(2|1,q_1^*)=P(1|2,q_1^*)$, 这就与每个贝叶斯判别都是容许的矛盾.

---

① 如果 $q_1=0$, 由 6.3.1 节中的 (5) 式知 $q_2p_2(\boldsymbol{x})\leqslant 0$, 从而 $p_2(\boldsymbol{x})=0,\boldsymbol{x}\in R_1$. —— 译者注

② $p_1(\boldsymbol{x})/p_2(\boldsymbol{x})=\infty$ 意味着 $p_2(\boldsymbol{x})=0$.

6.4 两多元正态总体的判别

现在我们把上面的判别方法应用于两个具有相同协方差阵的多元正态总体，即 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 和 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$，其中 $\boldsymbol{\mu}^{(i)\prime} = (\mu_1^{(i)}, \cdots, \mu_p^{(i)})$ 是第 i 个总体的均值向量，$i = 1, 2$，$\boldsymbol{\Sigma}$ 是每个总体的协方差阵 [Wald (1944) 第一次使用了这种方法]. 则第 i 个总体的密度为

$$p_i(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})\right]. \tag{1}$$

密度之比为

$$\begin{aligned}\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})} &= \frac{\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})\right]}{\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})\right]} \\ &= \exp\left\{-\frac{1}{2}[(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) - (\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})]\right\}.\end{aligned} \tag{2}$$

使得 (2) 式的值大于或等于 k(某个适当的 k) 的 x 组成的集合，就是判别为来自总体 π_1 对应的区域 R_1. 由于对数函数是递增的，从而上述不等式可以改写为

$$-\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)}) - (\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})\right] \geqslant \ln k. \tag{3}$$

(3) 式的左边可以展开为

$$\begin{aligned}&-\frac{1}{2}[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(1)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{\mu}^{(1)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{(1)} \\ &-\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{(2)} + \boldsymbol{\mu}^{(2)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{\mu}^{(2)\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^{(2)}].\end{aligned} \tag{4}$$

整理后可得

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}). \tag{5}$$

第一项就是著名的*判别函数*. 它是观测向量的函数.

下面的定理是定理 6.3.1 的直接推论.

定理 6.4.1 *若 π_i 的密度为 (1), $i = 1, 2$, 则最佳分类区域为*

$$\begin{aligned}R_1 &: \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \tfrac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \geqslant \ln k, \\ R_2 &: \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \tfrac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) < \ln k.\end{aligned} \tag{6}$$

若先验概率 q_1, q_2 是已知的，则 k 可以通过下式求出

$$k = \frac{q_2 C(1|2)}{q_1 C(2|1)}. \tag{7}$$

特别地，当两总体的先验概率与损失都相同时，$k = 1, \ln k = 0$. 此时总体 π_1 对应的分类区域为

$$R_1: \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \geqslant \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}). \tag{8}$$

若我们不知道先验概率, 则可以假设两种误判损失是相同的, 从而选择 $\ln k = c$. 设 $\boldsymbol{X}$ 是随机观测. 则我们希望分别在假设 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 和 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 的条件下求出下式的分布:

$$\boldsymbol{U} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}). \tag{9}$$

当 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 时 $\boldsymbol{U}$ 服从正态分布, 其均值与方差分别为

$$\begin{aligned} \mathrm{E}_1\boldsymbol{U} &= \boldsymbol{\mu}^{(1)\prime}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \\ &= \frac{1}{2}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \qquad (10) \\ \mathrm{Var}_1(\boldsymbol{U}) &= \mathrm{E}_1(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}^{(1)})(\boldsymbol{X}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \\ &= (\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}). \qquad (11) \end{aligned}$$

$N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 与 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 之间的 Mahalanobis (马哈拉诺比斯) 距离 (以下简称马氏距离) 的平方定义为

$$(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) = \Delta^2 \tag{12}$$

从而若 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$, 则 $\boldsymbol{U} \sim N(\frac{1}{2}\Delta^2, \Delta^2)$. 如果 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$, 则

$$\begin{aligned} \mathrm{E}_2\boldsymbol{U} &= \boldsymbol{\mu}^{(2)\prime}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) - \frac{1}{2}(\boldsymbol{\mu}^{(1)}+\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \\ &= \frac{1}{2}(\boldsymbol{\mu}^{(2)}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \qquad (13) \\ &= -\frac{1}{2}\Delta^2 \end{aligned}$$

由于 $\boldsymbol{U}$ 的方差只依赖于 $\boldsymbol{X}$ 的二阶矩, 所以 $\boldsymbol{U}$ 的方差与 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 时的方差相同. 从而 $\boldsymbol{U} \sim N(-\frac{1}{2}\Delta^2, \Delta^2)$.

若观测来自总体 π_1, 则误判概率为

$$P(2|1) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}\Delta} \mathrm{e}^{-\frac{1}{2}(z-\frac{1}{2}\Delta^2)^2/\Delta^2} \mathrm{d}z = \int_{-\infty}^{(c-\frac{1}{2}\Delta^2)/\Delta} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y; \tag{14}$$

若观测来自 π_2, 则误判概率为

$$P(1|2) = \int_{c}^{\infty} \frac{1}{\sqrt{2\pi}\Delta} \mathrm{e}^{-\frac{1}{2}(z+\frac{1}{2}\Delta^2)^2/\Delta^2} \mathrm{d}z = \int_{(c+\frac{1}{2}\Delta^2)/\Delta}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y. \tag{15}$$

图 6.1 用阴影部分在尾部表示出了这两种概率. 我们选择 c 使得

$$C(1|2) \int_{(c+\frac{1}{2}\Delta^2)/\Delta}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y = C(2|1) \int_{-\infty}^{(c-\frac{1}{2}\Delta^2)/\Delta} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y, \tag{16}$$

可求得极小化极大解.

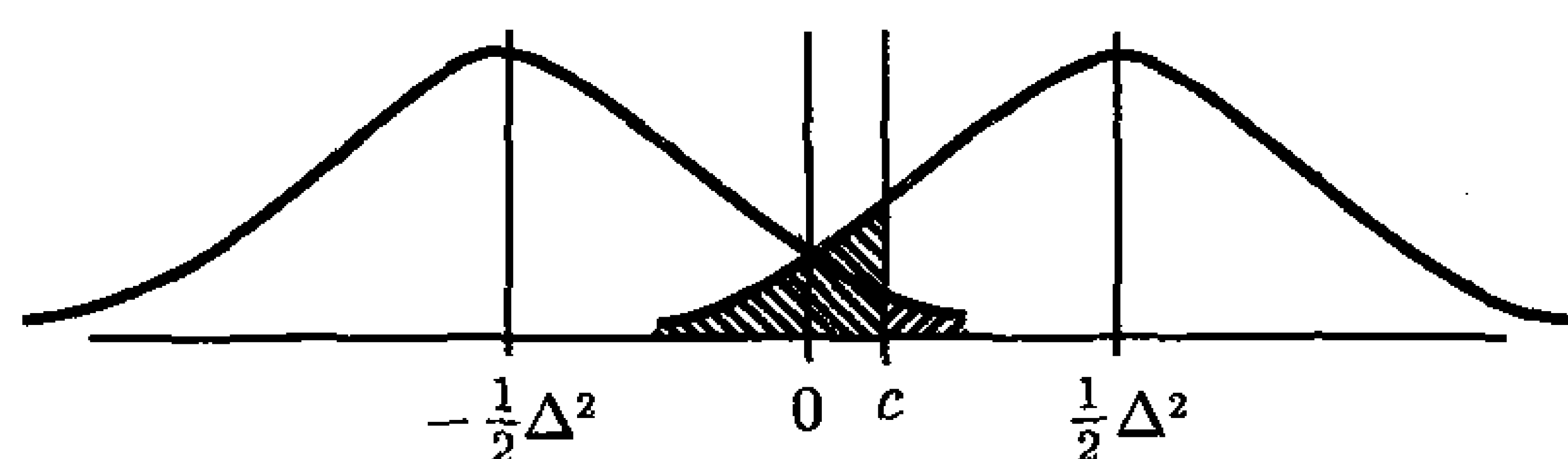


图 6.1

定理 6.4.2 如果 π_i 的密度为 (1), $i=1,2$, 则极小化极大分类区域由 (6) 式给出. 其中 $c=\ln k$ 由 16 式确定, $C(i|j)$ 表示两种误判损失.

注意, 当误判损失相同时, $c=0$, 误判概率为

$$\int_{\frac{\Delta}{2}}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}y^2}\mathrm{d}y. \tag{17}$$

当误判损失不同时, 我们可以利用试错法, 通过正态表把 c 足够精确地确定出来.

(5) 式中的两项都含有向量

$$\boldsymbol{\delta}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}). \tag{18}$$

这个向量可以通过一种有效的算法求解

$$\boldsymbol{\Sigma}\boldsymbol{\delta}=(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}) \tag{19}$$

而得. 对于任意的 $\boldsymbol{d}$, 线性判别函数 $\boldsymbol{x}'\boldsymbol{\delta}$ 极大化

$$\frac{[\mathrm{E}_1(\boldsymbol{X}'\boldsymbol{d})-\mathrm{E}_2(\boldsymbol{X}'\boldsymbol{d})]^2}{\mathrm{Var}(\boldsymbol{X}'\boldsymbol{d})}. \tag{20}$$

(20) 式的分子为

$$[\boldsymbol{\mu}^{(1)'}d-\boldsymbol{\mu}^{(2)'}d]^2=\boldsymbol{d}'[(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})']\boldsymbol{d}, \tag{21}$$

分母为

$$\boldsymbol{d}'\mathrm{E}(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X}))'\boldsymbol{d}=\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}. \tag{22}$$

我们希望在 (22) 式的限制条件下去求 $\boldsymbol{d}$, 使 (21) 式达到极大值. 设 λ 为拉格朗日乘子, 则只需极大化

$$\boldsymbol{d}'[(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})']\boldsymbol{d}-\lambda(\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}-1). \tag{23}$$

在 (23) 式两端对 $\boldsymbol{d}$ 的每一分量求导并令其等于 0 可得

$$2[(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})']\boldsymbol{d}=2\lambda\boldsymbol{\Sigma}\boldsymbol{d}. \tag{24}$$

因为 $(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{d}$ 是标量, 不妨设为 $\boldsymbol{\nu}$, (24) 式可以变形为

$$\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}=\frac{\lambda}{\boldsymbol{\nu}}\boldsymbol{\Sigma}\boldsymbol{d}. \tag{25}$$

因此上式的解与 $\boldsymbol{\delta}$ 成比例.

最后需要说明的是, 当有 N 个样本来自于 π_1 或 π_2 时, 由上面的讨论知, 可以利用样本均值把它们归入 $N(\boldsymbol{\mu}^{(1)},\frac{1}{N}\boldsymbol{\Sigma})$ 或者 $N(\boldsymbol{\mu}^{(2)},\frac{1}{N}\boldsymbol{\Sigma})$.

6.5　具有估计参数的两多元正态总体的判别

6.5.1　判别 (分类) 准则

在前面的讨论中, 我们总是假设两个总体是确切已知的. 但是在多数现实情况中总体是未知的, 我们必须通过从各自总体抽得的样本来推断它们的分布. 假设现在有两个正态总体以及各自的一组样本, 我们想利用这两组样本的信息, 判断另外一个观测来自这两个正态总体中的哪一个.

假设我们已有两组样本, 其中一组 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$, 另外一组 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_2}^{(1)}$ 来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$. 这些样本也被称为训练样本. 基于这些信息, 我们希望判断出观测 $\boldsymbol{x}$ 是属于 π_1 还是 π_2. 显然, $\boldsymbol{\mu}^{(1)}$ 与 $\boldsymbol{\mu}^{(2)}$ 的最优估计分别为 $\bar{\boldsymbol{x}}^{(1)} = \sum_1^{N_1} \boldsymbol{x}_\alpha^{(1)}/N_1$, $\bar{\boldsymbol{x}}^{(2)} = \sum_1^{N_2} \boldsymbol{x}_\alpha^{(2)}/N_2$. $\boldsymbol{\Sigma}$ 的估计为下面定义的 $\boldsymbol{S}$:

$$(N_1 + N_2 - 2)\boldsymbol{S} = \sum_{\alpha=1}^{N_1} (\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})' + \sum_{\alpha=1}^{N_2} (\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})'. \tag{1}$$

用这些估计代替 6.4 节 (5) 式中对应的参数就得到

$$\boldsymbol{W}(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}) - \frac{1}{2}(\bar{\boldsymbol{x}}^{(1)} + \bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}). \tag{2}$$

(2) 式的第一项是基于两组样本而得到的判别函数 [由 Fisher (1936) 提出]. 相对于*样本内方差*, 它是具有*最大样本间方差*的线性函数 (习题 6.12). 类似于 6.4 节 (5) 式的作用, (2) 式可以用来对样本进行分类.

在总体已知的前提下, 当先验概率已知时, 极小化期望损失的分类准则是最优的, 当先验概率未知时, 能够生成可容许判别程式类的分类准则是最优的. 但是, 我们无法由 (2) 式得到相同的结论. 然而, 凭直觉我们认为 (2) 式应该能够给出很好的结果. 6.5.5 节将介绍另外一个准则.

假设一组样本 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, 来自 π_1 或 π_2, 我们希望把这组样本整个地归入某个总体. 定义 $\boldsymbol{S}$ 为

$$(N_1 + N_2 + N - 3)\boldsymbol{S} = \sum_{\alpha=1}^{N_1} (\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})' + \sum_{\alpha=1}^{N_2} (\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})' + \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})', \tag{3}$$

其中

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha. \tag{4}$$

那么判别准则为

$$\left[\bar{\boldsymbol{x}} - \frac{1}{2}(\bar{\boldsymbol{x}}^{(1)} + \bar{\boldsymbol{x}}^{(2)})\right]' \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}). \tag{5}$$

N 越大, 误判概率越小.

6.5.2 判别准则的分布

对于随机的 $\boldsymbol{X}, \bar{\boldsymbol{X}}^{(1)}, \bar{\boldsymbol{X}}^{(2)}$ 和 $\boldsymbol{S}$, 假设

$$\begin{aligned}\boldsymbol{W} &= \boldsymbol{X}'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}) - \frac{1}{2}(\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}) \\ &= \left[\boldsymbol{X} - \frac{1}{2}(\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)})\right]' \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}).\end{aligned} \tag{6}$$

$\boldsymbol{W}$ 的分布非常复杂. 它依赖于样本容量和未知的 $\boldsymbol{\Delta}^2$. 设

$$\boldsymbol{Y}_1 = c_1[\boldsymbol{X} - (N_1+N_2)^{-1}(N_1\bar{\boldsymbol{X}}^{(1)} + N_2\bar{\boldsymbol{X}}^{(2)})], \tag{7}$$

$$\boldsymbol{Y}_2 = c_2(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}), \tag{8}$$

其中 $c_1 = \sqrt{(N_1+N_2)/(N_1+N_2+1)}$, $c_2 = \sqrt{N_1N_2/(N_1+N_2)}$. 则 $\boldsymbol{Y}_1$ 与 $\boldsymbol{Y}_2$ 相互独立, 都服从正态分布, 并且具有相同的协方差阵 $\boldsymbol{\Sigma}$. $\boldsymbol{Y}_2$ 的期望为 $c_2(\boldsymbol{\mu}^{(1)}\boldsymbol{\mu}^{(2)})$, 当 $\boldsymbol{X}$ 来自 π_1 时 $\boldsymbol{Y}_1$ 的期望为 $c_1[N_2/(N_1+N_2)](\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$, 当 $\boldsymbol{X}$ 来自于 π_2 时 $\boldsymbol{Y}_1$ 的期望为 $-c_1[N_1/(N_1+N_2)](\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$. 设 $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$,

$$\boldsymbol{M} = \boldsymbol{Y}'\boldsymbol{S}^{-1}\boldsymbol{Y} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22}\end{pmatrix}. \tag{9}$$

则

$$\boldsymbol{W} = \sqrt{\frac{N_1+N_2+1}{N_1N_2}}m_{12} + \frac{N_1-N_2}{2N_1N_2}m_{22}. \tag{10}$$

Sitgreaves(1952) 给出了 $\boldsymbol{M}$ 的密度. Anderson(1951a) 和 Wald(1944) 也讨论了 $\boldsymbol{M}$ 的分布.

如果 $N_1 = N_2$, 则当 $\boldsymbol{X}$ 来自 π_1 时 $\boldsymbol{W}$ 的分布与当 $\boldsymbol{X}$ 来自于 π_2 时 $-\boldsymbol{W}$ 的分布相同. 因此, 如果 $\boldsymbol{W} \geqslant 0$ 是判别为来自 π_1 的分类区域, 则 $\boldsymbol{X}$ 来自 π_1 的误判概率等于 $\boldsymbol{X}$ 来自 π_2 的误判概率.

6.5.3 准则的渐近正态分布

当来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 和 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 的样本比较多时, 我们可以利用极限分布理论. 因为 $\bar{\boldsymbol{X}}^{(1)}$ 是来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 的 N_1 个独立观测的均值, 从而

$$\operatorname*{plim}_{N_1\to\infty} \bar{\boldsymbol{X}}^{(1)} = \boldsymbol{\mu}^{(1)}. \tag{11}$$

(11) 式意味着: 给定任意的正数 δ 和 ε, 可以找到足够大的 N, 使得当 $N_1 \geqslant N$ 时有

$$\Pr\{|\bar{\boldsymbol{X}}_i^{(1)} - \boldsymbol{\mu}_i^{(1)}| < \delta, i = 1, \cdots, p\} > 1 - \varepsilon. \tag{12}$$

(见习题 3.23). 使用 Tchebycheff 不等式可以证明该结论. 同理

$$\operatorname*{plim}_{N_2 \to \infty} \bar{\boldsymbol{X}}^{(2)} = \boldsymbol{\mu}^{(2)}, \tag{13}$$

当 $N_1 \to \infty$ 或者 $N_2 \to \infty$ 或者两者同时趋于无穷时, 有

$$\operatorname{plim} \boldsymbol{S} = \boldsymbol{\Sigma}. \tag{14}$$

由上式可知

$$\operatorname{plim} \boldsymbol{S}^{-1} = \boldsymbol{\Sigma}^{-1}, \tag{15}$$

由于只要分母的概率极限不为零, 则随机变量和、差、积、商的概率极限等于随机变量概率极限的和、差、积、商 [Cramér, p254]. 而且

$$\operatorname*{plim}_{N_1, N_2 \to \infty} \boldsymbol{S}^{-1}(\bar{X}^{(1)} - \bar{\boldsymbol{X}}^{(2)}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}), \tag{16}$$

$$\operatorname*{plim}_{N_1, N_2 \to \infty} (\bar{\boldsymbol{X}}^{(1)} + \bar{\boldsymbol{X}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}^{(2)}) = (\boldsymbol{\mu}^{(1)} + \boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)}). \tag{17}$$

于是, $\boldsymbol{W}$ 的极限分布就是 $\boldsymbol{U}$ 的分布. 从而当来自 π_1, π_2 的样本量足够大时, 我们就可以利用分布完全已知时的准则进行判别, 并且判别误差非常小. [这个结果由 Wald (1944) 首次给出.]

定理 6.5.1 设 $\bar{\boldsymbol{X}}^{(1)}$ 是来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 的 N_1 个样本的均值, $\bar{\boldsymbol{X}}^{(2)}$ 是来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 的 N_2 个样本的均值, $\boldsymbol{S}$ 是基于总样本对 $\boldsymbol{\Sigma}$ 的一个估计, $\boldsymbol{W}$ 由 (6) 式给出. 当 $N_1 \to \infty$ 和 $N_2 \to \infty$ 时, 若 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma})$ 则 $\boldsymbol{W}$ 的极限分布为 $N(\frac{1}{2}\boldsymbol{\Delta}^2, \boldsymbol{\Delta}^2)$; 若 $\boldsymbol{X} \sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 则 $\boldsymbol{W}$ 的极限分布为 $N(-\frac{1}{2}\boldsymbol{\Delta}^2, \boldsymbol{\Delta}^2)$

6.5.4 另外一种判别准则

利用哑变量回归可以方便地推导判别准则 [Fisher (1936)]. 设

$$y_\alpha^{(1)} = \frac{N_2}{N_1 + N_2}, \alpha = 1, \cdots, N_1, \quad y_\alpha^{(2)} = \frac{-N_1}{N_1 + N_2}, \alpha = 1, \cdots, N_2. \tag{18}$$

对变量 $\boldsymbol{x}_\alpha^{(i)}$ 求回归, 即选择使下式极小化的 $\boldsymbol{b}$:

$$\sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} [y_\alpha^{(i)} - \boldsymbol{b}'(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}})]^2, \tag{19}$$

其中

$$\bar{\boldsymbol{x}} = \frac{N_1 \bar{\boldsymbol{x}}^{(1)} + N_2 \bar{\boldsymbol{x}}^{(2)}}{N_1 + N_2}. \tag{20}$$

正规方程为

$$\sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}})'\boldsymbol{b} = \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} y_\alpha^{(i)}(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}) \tag{21}$$

$$= \frac{N_1N_2}{N_1+N_2}[(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}) - (\bar{\boldsymbol{x}}^{(2)} - \bar{\boldsymbol{x}})]$$
$$= \frac{N_1N_2}{N_1+N_2}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}).$$

乘 $\boldsymbol{b}$ 的矩阵可以写为

$$\begin{aligned}
&\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}})' \\
=&\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})' \\
&+N_1(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}})' + N_2(\bar{\boldsymbol{x}}^{(2)} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(2)} - \bar{\boldsymbol{x}})' \\
=&\sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})' \\
&+\frac{N_1N_2}{N_1+N_2}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'.
\end{aligned} \tag{22}$$

因此 (21) 式可以写为

$$\boldsymbol{A}\boldsymbol{b} = (\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})\left[\frac{N_1N_2}{N_1+N_2} - \frac{N_1N_2}{N_1+N_2}\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'\boldsymbol{b}\right], \tag{23}$$

其中

$$\boldsymbol{A} = \sum_{i=1}^{2}\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})'. \tag{24}$$

因为 $(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'\boldsymbol{b}$ 是个标量, 所以 (23) 式的解 $\boldsymbol{b}$ 与 $\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})$ 成比例.

6.5.5 似然比准则

另一个常用的分类标准是似然比准则. 考虑检验复合原假设: $\boldsymbol{x}, \boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}), \boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{N_2}^{(2)}$ 来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$; 对立的复合备择假设: $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}), \boldsymbol{x}, \boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{N_2}^{(2)}$ 来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$, 其中 $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}$ 未知. 在第一个假设下 $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}$ 的极大似然估计分别为

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_1^{(1)} =& \frac{N_1\bar{\boldsymbol{x}}^{(1)} + \boldsymbol{x}}{N_1+1}, \\
\hat{\boldsymbol{\mu}}_1^{(2)} =& \bar{\boldsymbol{x}}^{(2)}, \\
\hat{\boldsymbol{\Sigma}}_1 =& \frac{1}{N_1+N_2+1}\left[\sum_{\alpha=1}^{N_1}(\boldsymbol{x}_\alpha^{(1)} - \hat{\boldsymbol{\mu}}_1^{(1)})(\boldsymbol{x}_\alpha^{(1)} - \hat{\boldsymbol{\mu}}_1^{(1)})' + (\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1^{(1)})(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1^{(1)})'\right. \\
&\left.+\sum_{\alpha=1}^{N_2}(\boldsymbol{x}_\alpha^{(2)} - \hat{\boldsymbol{\mu}}_1^{(2)})(\boldsymbol{x}_\alpha^{(2)} - \hat{\boldsymbol{\mu}}_1^{(2)})'\right],
\end{aligned} \tag{25}$$

因为

$$\begin{aligned}
&\sum_{\alpha=1}^{N_1}(\boldsymbol{x}_\alpha^{(1)}-\hat{\boldsymbol{\mu}}_1^{(1)})(\boldsymbol{x}_\alpha^{(1)}-\hat{\boldsymbol{\mu}}_1^{(1)})'+(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1^{(1)})(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1^{(1)})' \qquad (26)\\
=&\sum_{\alpha=1}^{N_1}(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})'+N_1(\bar{\boldsymbol{x}}^{(1)}-\hat{\boldsymbol{\mu}}_1^{(1)})(\bar{\boldsymbol{x}}^{(1)}-\hat{\boldsymbol{\mu}}_1^{(1)})'\\
&+(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1^{(1)})(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1^{(1)})'\\
=&\sum_{\alpha=1}^{N_1}(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})'+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})',
\end{aligned}$$

所以我们可以把 $\hat{\boldsymbol{\Sigma}}_1$ 写成

$$\hat{\boldsymbol{\Sigma}}_1=\frac{1}{N_1+N_2+1}\left[\boldsymbol{A}+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\right], \tag{27}$$

其中 $\boldsymbol{A}$ 由 (24) 式给出. 在备择假设下我们得到各个参数的极大似然估计为 (利用对称性)

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_2^{(1)}&=\bar{\boldsymbol{x}}^{(1)},\\
\hat{\boldsymbol{\mu}}_2^{(2)}&=\frac{N_2\bar{\boldsymbol{x}}^{(2)}+\boldsymbol{x}}{N_2+1}, \qquad (28)\\
\hat{\boldsymbol{\Sigma}}_2&=\frac{1}{N_1+N_2+1}\left[\boldsymbol{A}+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\right].
\end{aligned}$$

因此似然比准则为

$$\frac{|\hat{\boldsymbol{\Sigma}}_2|}{|\hat{\boldsymbol{\Sigma}}_1|}=\frac{|\boldsymbol{A}+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'|}{|\boldsymbol{A}+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'|} \tag{29}$$

的 $(N_1+N_2+1)/2$ 次幂. 上式可化为 (推论 A.3.1)

$$\begin{aligned}
&\frac{1+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{A}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})}{1+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{A}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})} \qquad (30)\\
=&\frac{n+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})}{n+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})},
\end{aligned}$$

其中 $n=N_1+N_2-2$. 那些使比式 (30) 大于或等于某个给定数 K_n 的点, 组成了对应于 π_1 的分类区域. 这个区域可以表示为

$$\begin{aligned}
R_1: n&+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)}) \qquad (31)\\
&\geqslant K_n\left[n+\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})\right].
\end{aligned}$$

如果 $K_n=1+2c/n$ 且 N_1 和 N_2 比较大, 区域 (31) 近似为 $\boldsymbol{W}(\boldsymbol{x})\geqslant c$.

如果 $K_n=1$, 则当 (30) 大于或等于 1 时我们把点归入 π_1, 当 (30) 小于 1 时我们把点归入 π_2. 这就是极大似然准则. 设

$$Z=\frac{1}{2}\left[\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})-\frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'S^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})\right]. \quad (32)$$

那么极大似然准则就是：若 $\boldsymbol{Z}>0$ 我们把样本归入 π_1，若 $\boldsymbol{Z}<0$ 就把样本归入 π_2. 粗略地说，$\boldsymbol{x}$ 是归入 π_1 还是 π_2 取决于 $\boldsymbol{x}$ 距离 $\bar{\boldsymbol{x}}^{(1)}$ 和 $\bar{\boldsymbol{x}}^{(2)}$ 哪一个更近. $\boldsymbol{W}$ 和 $\boldsymbol{Z}$ 之间的差为

$$\begin{aligned}\boldsymbol{W}-\boldsymbol{Z}=\frac{1}{2}\Bigg[&\frac{1}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})\\&-\frac{1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})\Bigg],\end{aligned} \quad (33)$$

当 $N_1,N_2\to\infty$ 时上式的概率极限为 0. 所以当样本量比较大时，$\boldsymbol{W}$ 和 $\boldsymbol{Z}$ 的误判概率比较接近.

注意，当 $N_1=N_2, \boldsymbol{Z}=[\frac{N_1}{N_1+1}]\boldsymbol{W}$ 时，在不考虑 $c=0$ 的情况下，$\boldsymbol{Z}$ 和 $\boldsymbol{W}$ 的对称检验是相同的.

6.5.6 不变性

对于变换

$$\begin{aligned}\boldsymbol{x}_\alpha^{(1)*}&=\boldsymbol{B}\boldsymbol{x}_\alpha^{(1)}+\boldsymbol{c},\quad \alpha=1,\cdots,N_1,\\ \boldsymbol{x}_\alpha^{(2)*}&=\boldsymbol{B}\boldsymbol{x}_\alpha^{(2)}+\boldsymbol{c},\quad \alpha=1,\cdots,N_2,\\ \boldsymbol{x}^*&=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{c},\end{aligned} \quad (34)$$

分类问题保持不变，其中 $\boldsymbol{B}$ 为非奇异矩阵，$\boldsymbol{c}$ 为向量. 这个变换归纳出基于充分统计量的变换：

$$\begin{aligned}&\bar{\boldsymbol{x}}^{(1)*}=\boldsymbol{B}\bar{\boldsymbol{x}}^{(1)}+\boldsymbol{c},\quad \bar{\boldsymbol{x}}^{(2)*}=\boldsymbol{B}\bar{\boldsymbol{x}}^{(2)}+\boldsymbol{c},\\ &\boldsymbol{x}^*=\boldsymbol{B}\boldsymbol{x}+\boldsymbol{c},\qquad \boldsymbol{S}^*=\boldsymbol{B}\boldsymbol{S}\boldsymbol{B}',\end{aligned} \quad (35)$$

对参数 $\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)}$ 和 $\boldsymbol{\Sigma}$ 必须做同样的变换 (注意 $\mathrm{E}(\boldsymbol{x})=\boldsymbol{\mu}^{(1)}$ 或 $\boldsymbol{\mu}^{(2)}$). 参数的任何不变量都是 $\boldsymbol{\Delta}^2=(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)})$ 的函数. 因为存在矩阵 $\boldsymbol{B}$ 和向量 $\boldsymbol{c}$，满足

$$\begin{aligned}&\boldsymbol{\mu}^{(1)*}=\boldsymbol{B}\boldsymbol{\mu}^{(1)}+\boldsymbol{c}=\boldsymbol{0},\quad \boldsymbol{\mu}^{(2)*}=\boldsymbol{B}\boldsymbol{\mu}^{(2)}+\boldsymbol{c}=(\boldsymbol{\Delta},0,\cdots,0)',\\ &\boldsymbol{\Sigma}^*=\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}'=\boldsymbol{I},\end{aligned} \quad (36)$$

所以 $\boldsymbol{\Delta}^2$ 是参数的极小不变量. 由 (9) 式定义的 $\boldsymbol{M}$ 的元素是不变量并且是充分统计量的极小不变量. 因此不变的判别过程依赖 $\boldsymbol{M}$，而 $\boldsymbol{M}$ 的分布只依赖 $\boldsymbol{\Delta}^2$. 统计量 $\boldsymbol{W}$ 和 $\boldsymbol{Z}$ 是不变量.

6.6 误判概率

6.6.1 利用 $\boldsymbol{W}$ 对误判概率进行渐近展开

在抽出两类样本进行分类之前和之后，我们希望知道误判概率或误判的 (条件)

概率. 在早期的观察中, $\boldsymbol{W}$ 和 $\boldsymbol{Z}$ 的准确分布难以计算. 因此, 随着 N_1, N_2 的增大, 我们把它们的概率进行渐近展开. 渐近展开的前提是若 $\boldsymbol{x}$ 来自 $\pi_1, \boldsymbol{W}$ 和 $\boldsymbol{Z}$ 的极限分布为 $N(\frac{1}{2}\Delta^2, \Delta^2)$, 若 $\boldsymbol{x}$ 来自于 $\pi_2, \boldsymbol{W}$ 和 $\boldsymbol{Z}$ 的极限分布为 $N(-\frac{1}{2}\Delta^2, \Delta^2)$.

Okamoto (1963) 得到 $\boldsymbol{W}$ 的分布的 n^{-2} 阶的渐近展开, Siotani and Wang (1975, 1977) 得到了 n^{-3} 阶的结果. [Bowker and Sitgreaves (1961) 讨论了 $N_1 = N_2$ 的情况.] 设 $\Phi(\cdot)$ 和 $\phi(\cdot)$ 分别为 $N(0,1)$ 的 cdf 和密度函数.

定理 6.6.1　当 $N_1 \to \infty, N_2 \to \infty, N_1/N_2 \to$ 一正极限时 $(n = N_1 + N_2 - 2)$,

$$\begin{aligned}&\Pr\left\{\frac{\boldsymbol{W}-\frac{1}{2}\boldsymbol{\Delta}^2}{\boldsymbol{\Delta}} \leqslant u \middle| \pi_1\right\} \qquad (1)\\
=&\Phi(u)-\phi(u)\left\{\frac{1}{2N_1\boldsymbol{\Delta}^2}[u^3+(p-3)u-p\boldsymbol{\Delta}]\right.\\
&+\frac{1}{2N_2\boldsymbol{\Delta}^2}[u^3+2\boldsymbol{\Delta}u^2+(p-3+\boldsymbol{\Delta}^2)u+(p-2)\boldsymbol{\Delta}]\\
&\left.+\frac{1}{4n}[4u^3+4\boldsymbol{\Delta}u^2+(6p-6+\boldsymbol{\Delta}^2)u+2(p-1)\boldsymbol{\Delta}]\right\}+O(n^{-2}),\end{aligned}$$

交换 (1) 式中的 N_1 和 N_2 就得到 $\Pr\left\{-\frac{\boldsymbol{W}+\frac{1}{2}\boldsymbol{\Delta}^2}{\boldsymbol{\Delta}} \leqslant u \middle| \pi_2\right\}$.

如果使用 $\boldsymbol{W}$ 作为观测 $\boldsymbol{x}$ 的判别准则, 那么当 $\boldsymbol{W}(x) > c$ 时, 把 $\boldsymbol{x}$ 归入 π_1; 当 $\boldsymbol{W}(x) \leqslant c$ 时, 把 $\boldsymbol{x}$ 归入 π_2. 定理 6.6.1 给出了这两种情况的误判概率 $u = (c-\frac{1}{2}\boldsymbol{\Delta}^2)/\boldsymbol{\Delta}$ 和 $u = -(c+\frac{1}{2}\boldsymbol{\Delta}^2)/\boldsymbol{\Delta}$. 如果 $c = 0, u = -\frac{1}{2}\boldsymbol{\Delta}$, 当 $N_1 = N_2$ 时, 这定义了一个精确的极小化极大程式 [Das Gupta (1965)].

推论 6.6.1

$$\begin{aligned}&\Pr\left\{\boldsymbol{W} \leqslant 0 \middle| \pi_1, \lim_{n\to\infty}\frac{N_1}{N_2}=1\right\} \qquad (2)\\
=&\Phi\left(-\frac{1}{2}\boldsymbol{\Delta}\right)+\frac{1}{n}\phi\left(\frac{1}{2}\boldsymbol{\Delta}\right)\left[\frac{p-1}{\boldsymbol{\Delta}}+\frac{p}{4}\boldsymbol{\Delta}\right]+o(n^{-1})\\
=&\Pr\left\{\boldsymbol{W} \geqslant 0 \middle| \pi_2, \lim_{n\to\infty}\frac{N_1}{N_2}=1\right\}.\end{aligned}$$

因为这个修正项大于 0, 所以误判概率要大于正态近似值. 给定 $\boldsymbol{\Delta}$, 修正项 (到 n^{-1} 阶) 随着 p 的增大而增大; 给定 p, 修正项 (到 n^{-1} 阶) 随着 $\boldsymbol{\Delta}$ 的增大而减小.

因为一般情况下 $\boldsymbol{\Delta}$ 未知, 所以 $\boldsymbol{\Delta}$ 和学生氏分布$\boldsymbol{W}$相关. 样本马氏距离的平方

$$\boldsymbol{D}^2 = (\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)}) \qquad (3)$$

是总体马氏距离平方 $\boldsymbol{\Delta}^2$ 的一个估计. $\boldsymbol{D}^2$ 的期望为

$$\mathrm{E}(\boldsymbol{D}^2) = \frac{n}{n-p-1}\left[\boldsymbol{\Delta}^2 + p\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\right]. \qquad (4)$$

见习题 6.14. 如果 N_1, N_2 比较大, 上式接近于 $\boldsymbol{\Delta}^2$.

Anderson(1973b) 给出了下面定理.

定理 6.6.2 如果 $n\to\infty$ 时 $N_1/N_2\to$ 一个正极限, 那么

$$\Pr\left\{\frac{\boldsymbol{W}-\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}}\leqslant u\middle|\pi_1\right\} \tag{5}$$

$$=\Phi(u)-\phi(u)\left\{\frac{1}{N_1}\left(\frac{u}{2}-\frac{p-1}{\boldsymbol{\Delta}}\right)+\frac{1}{n}\left[\frac{u^3}{4}+(p-\frac{3}{4})u\right]\right\}+\boldsymbol{O}(n^{-2}),$$

$$\Pr\left\{-\frac{\boldsymbol{W}+\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}}\leqslant u\middle|\pi_2\right\} \tag{6}$$

$$=\Phi(u)-\phi(u)\left\{\frac{1}{N_2}\left(\frac{u}{2}-\frac{p-1}{\boldsymbol{\Delta}}\right)+\frac{1}{n}\left[\frac{u^3}{4}+\left(p-\frac{3}{4}\right)u\right]\right\}+\boldsymbol{O}(n^{-2}).$$

一般来说, 我们只对 $u\leqslant 0$ 的情况感兴趣 (小的误差概率). 这时修正项是正的, 因此正态近似值低估了误判概率.

有时我们希望通过选择合适的截点 c 来控制误判概率. 设 α 是一个要求的 $\Pr\{\boldsymbol{W}<c|\pi_1\}$. Anderson (1973b, 1973c) 给出了下面的定理.

定理 6.6.3 设 u_0 满足 $\Phi(u_0)=\alpha$ 且

$$u=u_0-\frac{1}{N_1}\left[\frac{p-1}{\boldsymbol{D}}-\frac{1}{2}u_0\right]+\frac{1}{n}\left[\left(p-\frac{3}{4}\right)u_0+\frac{1}{4}u_0^3\right], \tag{7}$$

则当 $N_1\to\infty, N_2\to\infty$ 以及 $N_1/N_2\to$ 一正极限时, 有

$$\Pr\left\{\frac{\boldsymbol{W}-\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}}\leqslant u\middle|\pi_1\right\}=\alpha+\boldsymbol{O}(n^{-2}). \tag{8}$$

这时 $c=\boldsymbol{D}u+\frac{1}{2}\boldsymbol{D}^2$ 在 $O(n^{-2})$ 的误差范围内达到所要求的概率 α.

下面我们来讨论当两组样本被抽定后误判概率的估计. 给定 $\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{S}$, 随机变量 $\boldsymbol{W}$ 服从正态分布, 当 $\boldsymbol{x}$ 来自于 $\pi_i, i=1,2$ 时, 条件均值为

$$\begin{aligned}\mathrm{E}(\boldsymbol{W}|\pi_i,\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S}&=\left[\boldsymbol{\mu}^{(i)}-\frac{1}{2}(\bar{\boldsymbol{x}}^{(1)}+\bar{\boldsymbol{x}}^{(2)})\right]'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})\\&=\boldsymbol{\mu}^{(i)}(\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S}),\end{aligned} \tag{9}$$

条件方差为

$$\begin{aligned}\mathrm{Var}(\boldsymbol{W}|\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S})&=(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})\\&=\boldsymbol{\sigma}^2(\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S}).\end{aligned} \tag{10}$$

我们可以看到上面的均值和方差都是样本的函数且其概率极限分别为

$$\begin{aligned}\operatorname*{plim}_{N_1,N_2\to\infty}\ \boldsymbol{\mu}^{(i)}(\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S})&=(-1)^{i-1}\tfrac{1}{2}\boldsymbol{\Delta}^2,\\\operatorname*{plim}_{N_1,N_2\to\infty}\ \boldsymbol{\sigma}^2(\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S})&=\boldsymbol{\Delta}^2.\end{aligned} \tag{11}$$

当 N_1, N_2 比较大时, 条件误判概率接近于极限正态概率 (相对于 $\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{S}$ 而言, 具有更高的概率).

当截点为 c 时, 在 $\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{S}$ 已知条件下的误判概率为

$$P(2|1, c, \bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}) = \Phi\left[\frac{c - \mu^{(1)}(\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})}{\sigma(\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})}\right], \tag{12}$$

$$P(1|2, c, \bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}) = 1 - \Phi\left[\frac{c - \mu^{(2)}(\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})}{\sigma(\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})}\right]. \tag{13}$$

把 (12) 式中的 c 写作 $Du_1+\frac{1}{2}D^2$, 则 $\Phi(\cdot)$ 的自变量为 $u_1 D/\sigma+(\bar{\boldsymbol{x}}^{(1)}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(1)}-\boldsymbol{\mu}^{(1)})/\sigma$, 当 $N_1 \to \infty, N_2 \to \infty$ 时第一项依概率收敛于 u_1, 第二项趋向于 0, 从而 (12) 式收敛于 $\Phi(u_1)$. 把 (13) 式中的 c 写作 $Du_2 - \frac{1}{2}D^2$, 则 $\Phi(\cdot)$ 的自变量为 $u_2 D/\sigma + (\bar{\boldsymbol{x}}^{(1)} - \bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(2)} - \boldsymbol{\mu}^{(2)})/\sigma$, 第一项依概率收敛于 u_2, 第二项趋向于 0, 从而 (13) 式收敛于 $1 - \Phi(u_2)$.

给定 $\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{S}$, (条件) 误判概率 (12) 和 (13) 是参数 $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}$ 和 $\boldsymbol{\Sigma}$ 的函数并且是可估的. 当 $c = 0$ 时, (12),(13) 式依概率收敛于 $\Phi(-\frac{1}{2}\Delta)$, 所以我们建议用 $\Phi(-\frac{1}{2}D)$ 作为 (12) 和 (13) 的一个估计. 一个好的估计为 $\Phi(-\frac{1}{2}\bar{D})$, 其中 $\bar{D}^2 = (n-p-1)D^2/n$, 它几乎为 Δ^2 的一个无偏估计 [见 (4)]. McLachlan (1973, 1974a, 1974b, 1974c) 给出了 (12) 式的一个估计, 它的偏差为 n^{-2} 阶. 这个估计为

$$\Phi\left(-\frac{1}{2}D\right) + \phi\left(\frac{1}{2}D\right)\left\{\frac{p-D}{N_1 D} + \frac{1}{32n}[-D^3 + 4(4p-1)D]\right\}. \tag{14}$$

[McLachlan 把 (14) 式展开到 n^{-1} 阶项.] 与 Lachenbruch and Mickey (1968) 一样, McLachlan 研究了这些估计和其他一些估计的性质.

现在考虑 $c = Du_1 + \frac{1}{2}D^2$ 时的 (12) 式, 可以通过选定 u_1 来控制给定条件 $\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}$ 下的 $P(2|1)$. 这个条件概率是 $\bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}$ 和 $\boldsymbol{S}$ 的函数, 且为一个随机变量, 其分布可以被近似. McLachlan 得到了下述结果.

定理 6.6.4　当 $N_1 \to \infty, N_2 \to \infty, N_1/N_2 \to$ 一正极限时

$$\Pr\left\{\sqrt{n}\frac{P(2|1, Du_1 + \frac{1}{2}D^2, \bar{\boldsymbol{x}}^{(1)}, \bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}) - \Phi(u_1)}{\phi(u_2)[\frac{1}{2}u_1^2 + n/N_1]^{\frac{1}{2}}} \leqslant x\right\}, \tag{15}$$

$$= \Phi\left[x - \frac{(p-1)n/N_1 - (p - \frac{3}{4} + n/N_1)u_1 - u_1^3/4}{\sqrt{n}[\frac{1}{2}u_1^2 + n/N_1]^{\frac{1}{2}}}\right] + O(n^{-2}).$$

McLachlan (1977) 给出了一个选择 u_1 的方法, 这样选择的 u_1 可以使误判概率小于事先设置的 δ, 且 δ 具有预先设置的置信水平 $1-\varepsilon$.

6.6.2　利用 Z 对误判概率进行渐近展开

我们现在转向讨论 6.5 节中 (32) 式定义的 Z. 结果和 W 的一样. Memon and

Okamoto (1971) 把 Z 的分布展开到了 n^{-2} 阶项, Siotani and Wang (1975), (1977) 把 Z 的分布展到了 n^{-3} 阶项.

定理 6.6.5 当 $N_1 \to \infty, N_2 \to \infty$, N_1/N_2 趋向于正极限时, 则

$$\begin{aligned}
&\Pr\left\{\frac{Z-\frac{1}{2}\boldsymbol{\Delta}^2}{\boldsymbol{\Delta}} \leqslant u\Big|\pi_1\right\} \\
&=\Phi(u)-\phi(u)\left\{\frac{1}{2N_1\boldsymbol{\Delta}^2}[u^3+\boldsymbol{\Delta}u^2+(p-3)u-\boldsymbol{\Delta}]\right. \\
&\quad+\frac{1}{2N_2\boldsymbol{\Delta}^2}[u^3+\boldsymbol{\Delta}u^2+(p-3-\boldsymbol{\Delta}^2)u-\boldsymbol{\Delta}^3-\boldsymbol{\Delta}] \\
&\quad\left.+\frac{1}{4n}[4u^3+4\boldsymbol{\Delta}u^2+(6p-6+\boldsymbol{\Delta}^2)u+2(p-1)\boldsymbol{\Delta}]\right\}+O(n^{-2}),
\end{aligned} \tag{16}$$

把 (16) 式中的 N_1 和 N_2 交换一下就得到 $\Pr\left\{\left(-\frac{Z+\frac{1}{2}\boldsymbol{\Delta}^2}{\boldsymbol{\Delta}}\right)\leqslant u\Big|\pi_2\right\}$.

当 $c=0$ 时, $u=-\frac{1}{2}\Delta$. 若 $N_1=N_2$, 则 Z 的准则与 W 的准则是相同的, 并且误判概率由 (2) 式给出.

Fujikoshi and Kanazawa(1976) 证明了下述结果.

定理 6.6.6

$$\begin{aligned}
&\Pr\left\{\frac{Z-\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}} \leqslant u\Big|\pi_1\right\} \\
&=\Phi(u)-\phi(u)\left\{\frac{1}{2N_1\boldsymbol{\Delta}}[u^2+\boldsymbol{\Delta}u-(p-1)]\right. \\
&\quad-\frac{1}{2N_2\boldsymbol{\Delta}}[u^2+2\boldsymbol{\Delta}u+p-1+\boldsymbol{\Delta}^2] \\
&\quad\left.+\frac{1}{4n}[u^3+(4p-3)u]\right\}+O(n^{-2}),
\end{aligned} \tag{17}$$

$$\begin{aligned}
&\Pr\left\{-\frac{Z+\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}} \leqslant u\Big|\pi_2\right\} \\
&=\Phi(u)-\phi(u)\left\{-\frac{1}{2N_1\boldsymbol{\Delta}}\left[u^2+2\boldsymbol{\Delta}u+p-1+\boldsymbol{\Delta}^2\right]\right. \\
&\quad\left.+\frac{1}{2N_2\boldsymbol{\Delta}}\left[u^2+\boldsymbol{\Delta}u-(p-1)\right]+\frac{1}{4n}\left[u^3+(4p-3)u\right]\right\}+\boldsymbol{O}(n^{-2}).
\end{aligned} \tag{18}$$

Kanazawa (1979) 得到了下面的定理.

定理 6.6.7 设 u_0 满足 $\Phi(u_0)=\alpha$, 并设

$$\begin{aligned}
u=u_0&+\frac{1}{2N_1D}[u_0^2+Du_0-(p-1)] \\
&-\frac{1}{2N_2D}[u_0^2+Du_0+(p-1)-D^2]
\end{aligned} \tag{19}$$

$$+\frac{1}{4n}[u_0^3+(4p-5)u_0].$$

则当 $N_1\to\infty, N_2\to\infty$, $N_1/N_2\to$ 一正极限时,

$$\Pr\left\{\frac{\boldsymbol{Z}-\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}}\leqslant u\right\}=\alpha+\boldsymbol{O}(n^{-2}). \tag{20}$$

下面我们考虑抽取样本后的误判概率. 这时 $\boldsymbol{Z}$ 的条件分布不再是正态的, 除非 $N_1=N_2$, 否则 $\boldsymbol{Z}$ 是 $\boldsymbol{x}$ 的二次型. 这样就不存在与 (12) 和 (13) 式等价的表达式了. Siotani(1980) 得到了如下结果.

定理 6.6.8　$N_1\to\infty, N_2\to\infty$, $N_1/N_2\to$ 一正极限时,

$$\Pr\left\{2\sqrt{\frac{N_1N_2}{N_1+N_2}}\frac{P(2|1,0,\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S})-\Phi(-\frac{1}{2}\boldsymbol{\Delta})}{\phi(\frac{1}{2}\boldsymbol{\Delta})}\leqslant x\right\} \tag{21}$$

$$=\Phi\left[x-2\sqrt{\frac{N_1N_2}{N_1+N_2}}\left\{\frac{1}{16N_1\boldsymbol{\Delta}}[4(p-1)-\boldsymbol{\Delta}^2]\right.\right.$$

$$\left.\left.+\frac{1}{16N_2}[4(p-1)+3\boldsymbol{\Delta}^2]-\frac{(p-1)\boldsymbol{\Delta}}{4n}\right\}\right]+\boldsymbol{O}(n^{-2}).$$

给定 $\boldsymbol{Z}$ 和置信区间, 我们可以得到 $P(2|1,0,\boldsymbol{D}u_1+\frac{1}{2}\boldsymbol{D}^2,\bar{\boldsymbol{x}}^{(1)},\bar{\boldsymbol{x}}^{(2)},\boldsymbol{S})$ 的类似表达式. 见 Siotani(1980).

6.7　多总体的分类

现在考虑把观测归入多个总体中某一个的分类问题. 我们把前面几节考虑的情形扩展到多总体的情况. 设 $\pi_1,\cdots,\pi_m$ 是 m 个总体且分别具有密度函数 $p_1(\boldsymbol{x}),\cdots,p_m(\boldsymbol{x})$. 我们希望把整个观测空间分成分成互不相交的 m 个区域 $R_1,\cdots,R_m$. 如果观测落入区域 R_i, 就认为这个观测来自 π_i. 把一个来自 π_i 的观测误判为来自 π_j 的损失设为 $C(j|i)$. 那么误判概率为

$$P(j|i,R)=\int_{R_j}p_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{1}$$

假设总体的先验概率分别为 $q_1,\cdots,q_m$. 则期望损失为

$$\sum_{i=1}^m q_i\left\{\sum_{\substack{j=1\\j\neq i}}^m C(j|i)P(j|i,R)\right\}. \tag{2}$$

我们想找到适当的 $R_1,\cdots,R_m$ 使得上式取极小值.

因为每个总体的先验概率已知, 从而给定向量 $\boldsymbol{x}$ 就可以定义观测来自某一总体的条件概率. 观测来自总体 π_i 的条件概率为

$$\frac{q_i p_i(\boldsymbol{x})}{\sum_{k=1}^{m} q_k p_k(\boldsymbol{x})}. \tag{3}$$

那么把观测分入 π_j 的期望损失为

$$\sum_{\substack{i=1\\i\neq j}}^{m} \frac{q_i p_i(\boldsymbol{x})}{\sum_{k=1}^{m} q_k p_k(\boldsymbol{x})} C(j|i). \tag{4}$$

如果选择使 (4) 式极小化的 j, 那么期望损失就在这一点达到极小, 即要在所有的 j 中选择使

$$\sum_{\substack{i=1\\i\neq j}}^{m} q_i p_i(\boldsymbol{x}) C(j|i) \tag{5}$$

达到极小的 j. (如果有两个指标同时达到极小, 则任选其一.) 这样就把点 $\boldsymbol{x}$ 归入到了某个 R_j. 将这个过程应用于每个点 $\boldsymbol{x}$, 就定义了分类区域 $R_1, \cdots, R_m$. 如果观测落入 R_j, 我们就把它分入总体 π_j.

定理 6.7.1 设总体 π_i 的密度为 $p_i(\boldsymbol{x})$, $i = 1, \cdots, m$, 并且观测来自该总体的先验概率为 q_i. 设将来自 π_i 的观测误判为来自 π_j 的损失为 $C(j|i)$, 分类区域为 $R_1, \cdots, R_m$. 如果

$$\sum_{\substack{i=1\\i\neq k}}^{m} q_i p_i(\boldsymbol{x}) C(k|i) < \sum_{\substack{i=1\\i\neq j}}^{m} q_i p_i(\boldsymbol{x}) C(j|i), \quad j = 1, \cdots, m, \quad j \neq k, \tag{6}$$

我们通过把 $\boldsymbol{x}$ 分入到 R_k 来定义极小化期望损失. [如果除掉 h 个指标, (6) 式所有的 $j(j \neq k)$ 都成立, 并且把 (6) 式中的不等号换成等号时这 h 个指标满足该式, 则该点可以被分入这 $h+1$ 个总体的任何一个.] 如果在总体 π_i(每一个 i) 的条件下, 对所有的 k 和 j, (6) 式的左右两边相等的概率都为零, 则极小化判别在除零测集之外的情况下是唯一的.

证明 下面我们来证明这个结论. 设

$$h_j(\boldsymbol{x}) = \sum_{\substack{i=1\\i\neq j}}^{m} q_i p_i(\boldsymbol{x}) C(j|i). \tag{7}$$

则判别程式 R 的期望损失是

$$\sum_{j=1}^{m} \int_{R_j} h_j(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int h(\boldsymbol{x}|R) \mathrm{d}\boldsymbol{x}, \tag{8}$$

其中 $h(\boldsymbol{x}|R) = h_j(\boldsymbol{x}), \boldsymbol{x} \in R_j$. 对于定理中所描述的贝叶斯判别 R^*, $h(\boldsymbol{x}|R)$ 是 $h(\boldsymbol{x}|R^*) = \min\limits_i h_i(\boldsymbol{x})$. 因此, 任意分类程式 R 的期望损失和 R^* 的期望损失之差为

$$\int [h(\boldsymbol{x}|R) - h(\boldsymbol{x}|R^*)] \mathrm{d}\boldsymbol{x} = \sum_j \int_{R_j} [h_j(\boldsymbol{x}) - \min h_i(\boldsymbol{x})] \mathrm{d}\boldsymbol{x} \geqslant 0. \tag{9}$$

当且仅当 $h_i(\boldsymbol{x}) = \min\limits_i h_i(\boldsymbol{x}) (\boldsymbol{x} \in R_j)$ 时等号在相差一个零测集的基础上成立. ■

如果对任意的 $i, j, i \neq j$, 有 $C(j|i) = 1$, 则在 R_k 内

$$\sum_{\substack{i=1\\i\neq k}}^{m} q_i p_i(\boldsymbol{x}) < \sum_{\substack{i=1\\i\neq j}}^{m} q_i p_i(\boldsymbol{x}), \quad j \neq k. \tag{10}$$

在 (10) 式两边同时减去 $\sum\limits_{\substack{i=1\\i\neq k,j}}^{m} q_i p_i(\boldsymbol{x})$, 可得

$$q_j p_j(\boldsymbol{x}) < q_k p_k(\boldsymbol{x}), \quad j \neq k. \tag{11}$$

这时, 如果指标 k 所对应的 $q_i p_i(\boldsymbol{x})$ 是极大值, 则 $\boldsymbol{x}$ 就落入 R_k 内, 也就是说 $\boldsymbol{x}$ 最有可能来自总体 π_k.

如果我们不知道先验概率, 就无法对分类程式定义一个无条件的期望损失. 这时我们可以定义在观测来自某一给定总体条件下的期望损失. 若观测来自 π_i 的条件期望损失为

$$\sum_{\substack{j=1\\j\neq i}}^{m} C(j|i) P(j|i, R) = r(i, R). \tag{12}$$

对两个分类程式 R, R^*, 若 $r(i, R) \leqslant r(i, R^*), i = 1, \cdots, m$, 则称分类程式 R *优于* R^*; 若至少有一个不等号是严格成立的, 则称分类程式 R *严格优于* R^*. 若不存在分类过程 R^* 严格优于 R, 则称分类程式 R 为*容许的*. 称一个由分类程式组成的集合是完全的, 若对于集合外的任一个分类程式 R, 都存在集合内的一个分类程式 R^* 严格优于 R.

下面我们将证明贝叶斯判别是容许的. 设 R 为贝叶斯判别, R^* 为另一程式. 因为 R 是贝叶斯的, 所以

$$\sum_{i=1}^{m} q_i r(i, R) \leqslant \sum_{i=1}^{m} q_i r(i, R^*). \tag{13}$$

设 $q_1 > 0, q_2 > 0, r(2, R^*) < r(2, R), r(i, R^*) \leqslant r(i, R), i = 3, \cdots, m$. 则

$$q_1[r(1, R) - r(1, R^*)] \leqslant \sum_{i=2}^{m} q_i[r(i, R^*) - r(i, R)] < 0, \tag{14}$$

且 $r(1, R) < r(1, R^*)$. 因此 R^* 不严格优于 R.

定理 6.7.2 *若 $q_i > 0, i = 1, \cdots, m$, 则贝叶斯判别是容许的.*

假设 $C(i|j) = 1, i \neq j, \Pr\{p_i(\boldsymbol{x}) = 0|\pi_j\} = 0$. 后一个条件意味着所有的 $p_i(\boldsymbol{x})$ 在同一个集合上取正值 (零测集除外). 设 $q_i = 0, i = 1, \cdots, t, q_i > 0, i = t+1, \cdots, m$. 则由 (11) 式可知 [即 $p_m(\boldsymbol{x}) = 0, \boldsymbol{x} \in R_i$], 贝叶斯解 $R_i (i = 1, \cdots, t)$ 是空集 (零测集除外). 由此可得, $r(i, R) = \sum_{j\neq i} P(j|i, R) = 1 - P(i|i, R) = 1, i = 1, \cdots, t$. 则 $(R_{t+1}, \cdots, R_m)$ 是关于 $p_{t+1}(\boldsymbol{x}), \cdots, p_m(\boldsymbol{x})$ 和 $q_{t+1}, \cdots, q_m$ 的分类问题的贝叶斯解. 由定理 6.7.2 可知, 不存在分类程式 R^*, 严格优于贝叶斯判别, 其中 $P(i|i, R^*) =$

$0, i=1,\cdots,t$. 考虑满足 R_1^* 包含一个正概率集的程式 R^*，即 $P(1|1,R^*)>0$. 因为 R^* 严格优于 R,

$$\begin{aligned} P(i|i,R) &= \int_{R_i} p_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &\leqslant P(i|i,R^*) = \int_{R_i^*} p_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \quad i=2,\cdots,m. \end{aligned} \tag{15}$$

此时, 若有分类程式 R^{**} 满足 $R_i^{**}(i=1,\cdots,t)$ 是空集, $R_i^{**}=R_i^*(i=t+1,\cdots,m-1)$, $R_m^{**}=R_m^*\cup R_1^*\cup\cdots\cup R_t^*$，则它的风险是

$$\begin{aligned} P(i|i,R^{**}) &= 0, \quad i=1,\cdots,t, \\ P(i|i,R^{**}) &= P(i|i,R^*) \geqslant P(i|i,R), \quad i=t+1,\cdots,m-1, \\ P(m|m,R^{**}) &> P(m|m,R^*) \geqslant P(m|m,R). \end{aligned} \tag{16}$$

对于 $m-t$ 个决策问题, $(R_{t+1}^{**},\cdots,R_m^{**})$ 严格优于 $(R_{t+1},\cdots,R_m)$, 这与上面的讨论矛盾.

定理 6.7.3 *若 $C(i|j)=1, i\neq j, \Pr\{p_i(\boldsymbol{x})=0|\pi_j\}=0$, 则贝叶斯过程是容许的.*

去掉条件后, 其逆也是正确的 (除去参数空间是有限的情况).

定理 6.7.4 *每一个容许的分类程式都是贝叶斯判别.*

此处略去这个定理的证明, 它就是 Ferguson(1967) 中 2.10 节的定理 1. 如果每个贝叶斯判别是唯一的 (对于给定的概率), 则由贝叶斯判别组成的类就是最小完全类.

当风险都相同时, 极小化极大分类程式就是贝叶斯判别.

Wald(1950), Blackwell and Girshick(1954), Ferguson(1967), De Groot(1970), Berger(1980b) 给出了统计决策程式的一般处理方法.

6.8 多个多元正态总体的分类

我们现在把 6.7 节中的结论应用到多元正态总体的情况. [见 von Mises (1945).] 假设总体具有相同的协方差阵, 不同的均值向量, π_i 的分布为 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})$. 密度由 6.4 节中的 (1) 式给出. 在一开始我们假设参数都是已知的. 对于已知先验概率的一般损失, 我们可以得到 6.7 节 (5) 式中的 m 个函数, 并把使第 j 个函数达到极小的点 $\boldsymbol{x}$ 组成的集合定义为分类区域 R_j.

在下面的讨论中, 假设误判损失都是相同的. 那么利用函数

$$u_{jk}(\boldsymbol{x}) = \ln\frac{p_j(\boldsymbol{x})}{p_k(\boldsymbol{x})} = \left[\boldsymbol{x}-\frac{1}{2}(\boldsymbol{\mu}^{(j)}+\boldsymbol{\mu}^{(k)})\right]'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(j)}-\boldsymbol{\mu}^{(k)}). \tag{1}$$

若先验概率已知, 区域 R_j 由满足下面条件的 $\boldsymbol{x}$ 组成,

$$R_j : u_{jk}(\boldsymbol{x}) > \ln \frac{q_k}{q_j}, \quad k = 1, \cdots, m, \ k \neq j. \tag{2}$$

定理 6.8.1 若从 $\pi_i = N(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma})$ 中抽取观测的先验概率为 q_i, $i = 1, \cdots, m$, 且误判损失都是相同的, 那么由 (2) 式定义的分类区域 $R_1, \cdots R_m$ 极小化期望损失, 其中 $u_{jk}(\boldsymbol{x})$ 由 (1) 式给出.

应该注意到, 每一个 $u_{jk}(\boldsymbol{x})$ 都是与第 j 个总体和第 k 个总体相关的分类函数, 而且 $u_{jk}(\boldsymbol{x}) = -u_{kj}(\boldsymbol{x})$. 因为它们都是线性函数, 所以 R_i 可以通过超平面来界定. 如果均值张成了 $m-1$ 维的超平面 (例如, 向量 $\boldsymbol{\mu}^{(i)}$ 是线性独立的且 $p \geqslant m-1$), 则 R_i 可以由 $m-1$ 个超平面来界定.

当先验概率未知时, 可以通过下面的不等式定义分类区域 R_j:

$$u_{jk}(\boldsymbol{x}) \geqslant c_j - c_k, \quad k = 1, \cdots, m, \ k \neq j. \tag{3}$$

常量 c_k 可以取非负数. 由这些分类区域得到的分类程式组成了容许类. 若这些常数使得 $P(i|i, R)$ 全部相等, 则我们得到的就是极小化极大分类程式.

下面来讨论如何计算正确分类的概率. 设 $\boldsymbol{X}$ 为一随机观测, 考虑随机变量

$$U_{ji} = \left[\boldsymbol{X} - \frac{1}{2}(\boldsymbol{\mu}^{(i)} + \boldsymbol{\mu}^{(j)})\right]' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(i)}). \tag{4}$$

这里 $U_{ji} = -U_{ij}$. 因此, 若均值张成一个 $m-1$ 维的超平面, 则需要 $m(m-1)/2$ 个分类函数. 若 $\boldsymbol{X}$ 来自 π_j, 则 U_{ji} 服从分布 $N(\frac{1}{2}\boldsymbol{\Delta}_{ji}^2, \boldsymbol{\Delta}_{ji}^2)$, 其中

$$\boldsymbol{\Delta}_{ji}^2 = (\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(i)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(i)}). \tag{5}$$

U_{ji} 与 U_{jk} 的协方差阵为

$$\boldsymbol{\Delta}_{jk,ji} = (\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(k)})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^{(j)} - \boldsymbol{\mu}^{(i)}). \tag{6}$$

利用下面的积分来确定常量 c_j:

$$P(j|j, R) = \int_{c_j - c_m}^{\infty} \cdots \int_{c_j - c_1}^{\infty} f_j \mathrm{d}u_{j1} \cdots \mathrm{d}u_{j,j-1} \mathrm{d}u_{j,j+1} \cdots \mathrm{d}u_{jm}, \tag{7}$$

其中 f_j 是 $U_{ji}(i = 1, \cdots, m, \ i \neq j)$ 的密度.

定理 6.8.2 若 π_i 是 $N(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma})$ 并且误判损失都相同, 则由 (3) 式定义的分类区域 $R_1, \cdots, R_m$, 极小化极大条件期望损失, 其中 $u_{jk}(\boldsymbol{x})$ 由 (1) 式给出. 选取常数 c_j 使得 (7) 式中的积分全部相等.

以 $m = 3$ 为例, 不失一般性, 取 $p = 2$, 因为当三个总体的均值是非共线时, 高维 p 的密度都可以投影到由三个总体均值决定的二维平面 (即我们可以把向量 $\boldsymbol{x}$ 转化成 u_{12}, u_{13}, 和另外 $p-2$ 个坐标, 其中后面这 $p-2$ 个分量与 u_{12}, u_{13} 分布独立且均值为零). 如图 6.2 所示, 三条射线决定了分类区域 R_j. 如果这个程式是极

小化极大程式, 在不存在不包含在任何区域内的一个三角且保持等式 $P(1|1,R)=P(2|2,R)=P(3|3,R)$ 的条件下, 我们无法移动 R_1 与 R_2 之间的直线使其接近 $(\mu_1^{(1)},\mu_2^{(1)})$, 移动 R_2 与 R_3 之间的直线使其接近 $(\mu_1^{(2)},\mu_2^{(2)})$ 以及移动 R_3 与 R_1 之间的直线使其接近 $(\mu_1^{(3)},\mu_2^{(3)})$. 因为区域必须占满整个空间, 所以直线必定相交于

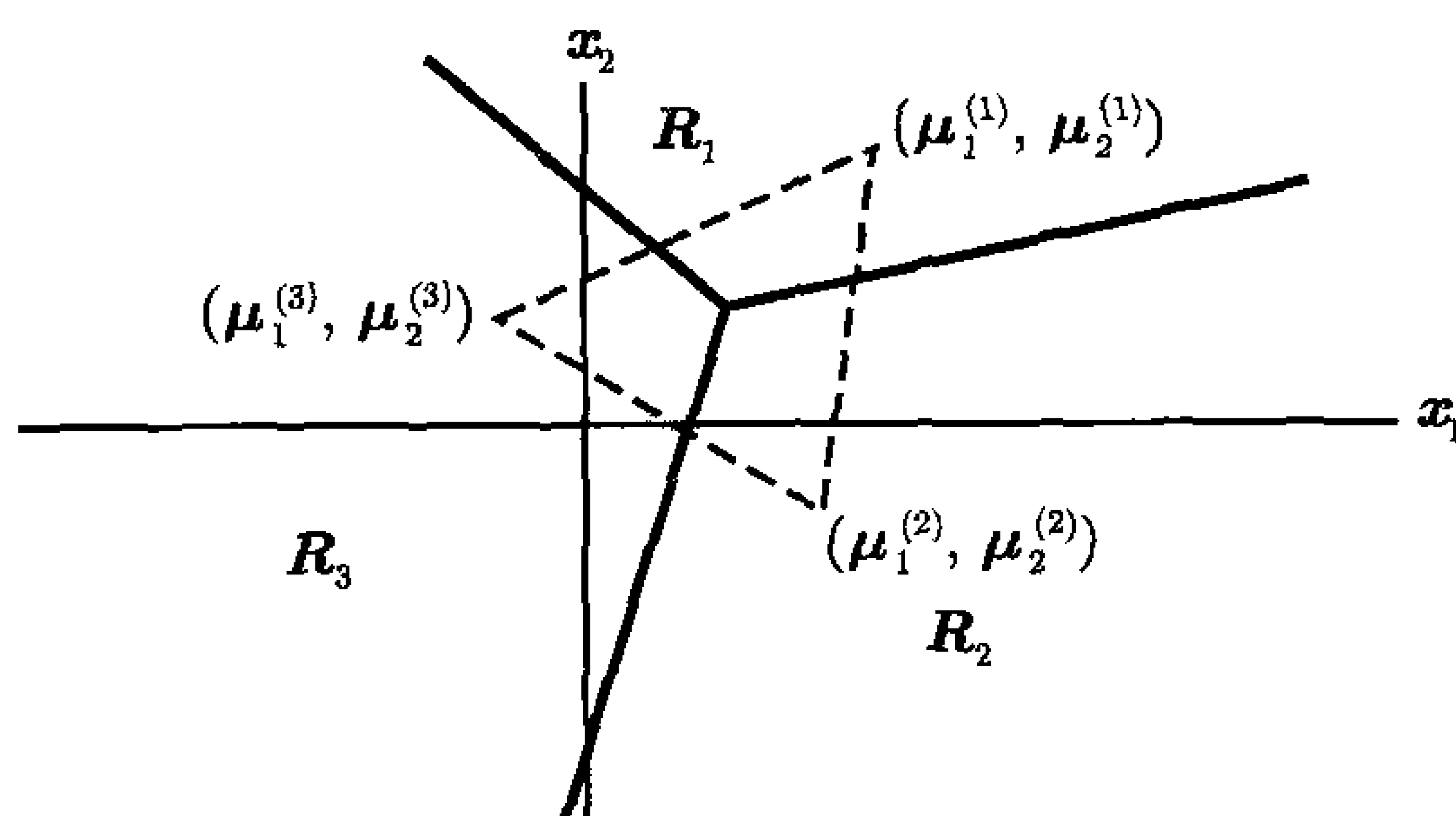


图 6.2 分类区域

某一点而且概率的等式唯一决定了 c_i-c_j.

为了具体说明这种情况, 假设我们已经知道向量 $\boldsymbol{\mu}^{(1)},\boldsymbol{\mu}^{(2)},\boldsymbol{\mu}^{(3)}$ 和矩阵 $\boldsymbol{\Sigma}$ 的各个分量的数值, 考虑三组 ($\leqslant p+1$) 联合分布, 每一组都有两个 $U_{ij}(j\neq i)$ 组成. 我们可以尝试取 $c_i=0$, 利用二元正态分布表 [Pearson(1931)] 计算 $P(i|i,R)$. 通过试错法得到 c_i 来近似上述情况.

上面的理论都是在参数已知的假设条件下得到的. 如果参数是未知的且每个总体的样本是可以得到的, 则可以用估计值代替 $u_{ij}(x)$ 定义中的参数. 假设观测 $\boldsymbol{x}_1^{(i)},\cdots,\boldsymbol{x}_{N_i}^{(i)}$ 来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}),i=1,\cdots,m$. 使用

$$\bar{\boldsymbol{x}}^{(i)}=\frac{1}{N_i}\sum_{\alpha=1}^{N_i}\boldsymbol{x}_\alpha^{(i)} \tag{8}$$

估计 $\boldsymbol{\mu}^{(i)}$, 利用 $\boldsymbol{S}$ 估计 $\boldsymbol{\Sigma}$, $\boldsymbol{S}$ 由下式定义:

$$\left(\sum_{i=1}^{m}N_i-m\right)\boldsymbol{S}=\sum_{i=1}^{m}\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})'. \tag{9}$$

则对应的 $u_{ij}(\boldsymbol{x})$ 为

$$w_{ij}(\boldsymbol{x})=\left[\boldsymbol{x}-\frac{1}{2}(\bar{\boldsymbol{x}}^{(i)}+\bar{\boldsymbol{x}}^{(j)})\right]'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}^{(i)}-\bar{\boldsymbol{x}}^{(j)}). \tag{10}$$

若上面的变量是随机的, 则它们的分布与 U_{ij} 的分布是不同的. 但是, 当 $N_i\to\infty$ 时, 联合分布接近于 U_{ij} 的分布. 因此, 当样本充分多时仍然可以利用上面的理论.

6.9 多个多元正态总体分类的一个例子

Rao (1948a) 考虑了三个总体, 分别为印度的婆罗门 (π_1), 技工 (π_2) 和 Korwa

人 (π_3). 我们测量各个阶层中每个个体的身高 (x_1), 坐高 (x_2), 鼻深 (x_3), 鼻高 (x_4). 三个总体对应的这些变量的均值都列在表 6.2 中. 总体的相关阵为

$$\begin{pmatrix} 1.0000 & 0.5849 & 0.1774 & 0.1974 \\ 0.5849 & 1.0000 & 0.2094 & 0.2170 \\ 0.1774 & 0.2094 & 1.0000 & 0.2910 \\ 0.1974 & 0.2170 & 0.2910 & 1.0000 \end{pmatrix} \tag{1}$$

表　6.2

测　量	均　值		
	婆罗门 (π_1)	技工 (π_2)	Korwa 人 (π_3)
身高 (x_1)	164.51	160.53	158.17
坐高 (x_2)	86.43	81.47	81.16
鼻深 (x_3)	25.49	23.84	21.44
鼻高 (x_4)	51.24	48.62	46.72

标准差分别为 $\sigma_1 = 5.74, \sigma_2 = 3.20, \sigma_3 = 1.75, \sigma_4 = 3.50$. 假设每个总体都是正态的. 我们的问题就是把四个变量 x_1, x_2, x_3, x_4 的空间分成三个分类区域. 假设误判损失都是相等的. 现在要找到分类区域使得: (i) 新观测来自每个总体的概率都相等 ($q_1 = q_2 = q_3 = \frac{1}{3}$), (ii) 最大的误判概率是极小化的 (极小化极大解).

首先计算 $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$ 与 $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(3)})$ 的系数. 则 $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(2)} - \boldsymbol{\mu}^{(3)}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(3)}) - \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})$. 接着计算 $\frac{1}{2}(\boldsymbol{\mu}^{(i)} + \boldsymbol{\mu}^{(j)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^{(i)} - \boldsymbol{\mu}^{(j)})$. 得到判别函数①

$$\begin{aligned} u_{12}(\bar{\boldsymbol{x}}) &= -0.0708x_1 + 0.4990x_2 + 0.3373x_3 + 0.0887x_4 - 43.13, \\ u_{13}(\bar{\boldsymbol{x}}) &= 0.0003x_1 + 0.3550x_2 + 1.1063x_3 + 0.1375x_4 - 62.49, \\ u_{23}(\bar{\boldsymbol{x}}) &= 0.0711x_1 - 0.1440x_2 + 0.7690x_3 + 0.0488x_4 - 19.36. \end{aligned} \tag{2}$$

其余的三个函数为 $u_{21}(\boldsymbol{x}) = -u_{12}(\boldsymbol{x}), u_{31}(\boldsymbol{x}) = -u_{13}(\boldsymbol{x})$ 和 $u_{32}(\boldsymbol{x}) = -u_{23}(\boldsymbol{x})$. 若先验概率已知且相等, 则最佳分类区域集为 $R_1 : u_{12}(\boldsymbol{x}) \geqslant 0, u_{13}(\boldsymbol{x}) \geqslant 0$; $R_2 : u_{21}(\boldsymbol{x}) \geqslant 0, u_{23}(\boldsymbol{x}) \geqslant 0$; $R_3 : u_{31}(\boldsymbol{x}) \geqslant 0, u_{32}(\boldsymbol{x}) \geqslant 0$. 例如, 如果一个个体所测得的指标 $\boldsymbol{x}$ 满足 $u_{12}(\boldsymbol{x}) \geqslant 0, u_{13}(\boldsymbol{x}) \geqslant 0$, 我们就认为他是婆罗门.

为了求得当个体来自总体 π_g 时的误判概率, 我们需要知道均值、方差、一些特定 u 对的协方差. 它们在表 6.3②中列出.

对于二元正态分布, 由表 6.3 就可得到误判概率. 对应于 π_1, π_2, π_3 的误判概率分别为 0.21, 0.42, 0.25. 比如, 通过测量把一个婆罗门误认为技工或 Korwa 人的概率为 0.21.

① 由于计算误差, Rao 的判别函数是不正确的. 我非常感谢 Peter Frank 先生帮助我们计算.

② 在 Anderson(1951a) 中有一些数字错误, 在表 6.3 和 (3) 中我们把它改正过来了.

表 6.3

x 的总体	u	均值	标准差	相关系数
π_1	u_{12}	1.491	1.727	0.8658
	u_{13}	3.487	2.641	
π_2	u_{21}	1.491	1.727	−0.3894
	u_{23}	1.031	1.436	
π_3	u_{31}	3.487	2.641	0.7983
	u_{32}	1.031	1.436	

在误判概率相等的条件下，通过求 6.3 节 (3) 式中的常量 c_1, c_2, c_3 可以得到极小化极大解. 分类区域为

$$
\begin{aligned}
&R_1': u_{12}(\boldsymbol{x}) \geqslant 0.54, \quad u_{13}(\boldsymbol{x}) \geqslant 0.29;\\
&R_2': u_{21}(\boldsymbol{x}) \geqslant -0.54, \quad u_{23}(\boldsymbol{x}) \geqslant -0.25;\\
&R_3': u_{31}(\boldsymbol{x}) \geqslant -0.29, \quad u_{32}(\boldsymbol{x}) \geqslant 0.25.
\end{aligned}
\tag{3}
$$

误判概率保留到小数点后两位是 0.30, 因此最大误判概率从 0.42 降到 0.30.

6.10 具有不同协方差阵的两多元正态总体的分类

6.10.1 似然比判别程式

设 π_1, π_2 的分布分别为 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1), N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$, 且 $\boldsymbol{\mu}^{(1)} \neq \boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. 当参数未知时, 似然比为

$$
\begin{aligned}
\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})} &= \frac{|\boldsymbol{\Sigma}_2|^{\frac{1}{2}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})]}{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})]} \qquad (1)\\
&= |\boldsymbol{\Sigma}_2|^{\frac{1}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{1}{2}} \exp\left[\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(2)}) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}^{(1)})\right].
\end{aligned}
$$

(1) 式的自然对数是 $\boldsymbol{x}$ 的二次型, 因此误判概率难以计算. [可以通过线性变换把 $\boldsymbol{x}$ 的协方差阵变为单位矩阵 I, 把二次型的矩阵变为对角的形式, 则 (1) 式的自然对数就变为一些非中心 χ^2 变量的线性组合再加上一个常量.]

当参数未知时, 我们把所考虑的问题转化为检验假设 $\boldsymbol{x}, \boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 是来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1)$ 的观测, $\boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{N_2}^{(2)}$ 是来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$ 的观测; 备择假设为 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 是来自 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1)$ 的观测, $\boldsymbol{x}, \boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{N_2}^{(2)}$ 是来自 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$ 的观测. 在第一个假设下, 极大似然估计分别为 $\hat{\boldsymbol{\mu}}_1^{(1)} = \frac{(N_1\bar{\boldsymbol{x}}^{(1)}+\boldsymbol{x})}{N_1+1}$, $\hat{\boldsymbol{\mu}}_1^{(2)} = \bar{\boldsymbol{x}}^{(2)}$,

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}_1(1) &= \frac{1}{N_1+1}\left[A_1 + \frac{N_1}{N_1+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\right],\\
\hat{\boldsymbol{\Sigma}}_2(1) &= \frac{1}{N_2}A_2,
\end{aligned}
\tag{2}
$$

其中 $A_i=\sum_{\alpha=1}^{N_i}(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})'$, $i=1,2$. (见 6.5.5 节.) 在第二个假设下, 极大似然估计分别为 $\hat{\boldsymbol{\mu}}_2^{(1)}=\bar{\boldsymbol{x}}^{(1)}$, $\hat{\boldsymbol{\mu}}_2^{(2)}=\frac{(N_2\bar{\boldsymbol{x}}^{(2)}+\boldsymbol{x})}{N_2+1}$,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_1(2)&=\frac{1}{N_1}A_1,\\ \hat{\boldsymbol{\Sigma}}_2(2)&=\frac{1}{N_2+1}\left[A_2+\frac{N_2}{N_2+1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\right].\end{aligned}\tag{3}$$

所以似然比准则为

$$\begin{aligned}\frac{|\hat{\boldsymbol{\Sigma}}_1(2)|^{\frac{1}{2}N_1}|\hat{\boldsymbol{\Sigma}}_2(2)|^{\frac{1}{2}(N_2+1)}}{|\hat{\boldsymbol{\Sigma}}_1(1)|^{\frac{1}{2}(N_1+1)}|\hat{\boldsymbol{\Sigma}}_2(1)|^{\frac{1}{2}N_2}}&=\frac{[1+(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{A}_2^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})]^{\frac{1}{2}(N_2+1)}}{[1+(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{A}_1^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})]^{\frac{1}{2}(N_1+1)}}\\ &\quad\cdot\frac{(N_1+1)^{\frac{1}{2}(N_1+1)p}N_2^{\frac{1}{2}N_2p}|\boldsymbol{A}_2|^{\frac{1}{2}}}{N_1^{\frac{1}{2}N_1p}(N_2+1)^{\frac{1}{2}(N_2+1)p}|\boldsymbol{A}_1|^{\frac{1}{2}}}.\end{aligned}\tag{4}$$

若 (4) 式大于 1, 我们就把观测 $\boldsymbol{x}$ 归入 π_1, 否则归入 π_2.

用估计值代替 (1) 的对数中的参数, 得到另外一种判别准则. 即

$$\frac{1}{2}[(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})'\boldsymbol{S}_2^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(2)})-(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})'\boldsymbol{S}_1^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}^{(1)})],\tag{5}$$

若 (5) 式比较大我们把观测 $\boldsymbol{x}$ 归入 π_1, 否则归入 π_2. 但误判概率仍然难以估计.

6.10.2 线性判别程式

当 $\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_2$ 时, 最佳判别程式是非线性的; 当参数已知时, 最佳判别是观测向量 $\boldsymbol{x}$ 的二次函数. 这个判别非常依赖于正态性假设. 例如, 当 $p=1$ 时, 某个总体的分类区域是一个区间, 其他总体的分类区域是这个区间的补集, 即观测充分大或充分小. 二变量的时候, 分类区域由圆锥曲线定义. 例如, 每个总体对应的分类区域可能为椭圆的内部或者两条双曲线之间的区域. 一般来说, 分类区域由观测值的二次函数决定而且这个二次型不一定是正定的. 这些判别过程都非常依赖于正态性假设, 尤其依赖于正态分布偏离中心的形状. 例如, 在上面所述的单变量情况, 第一个总体对应的分类区域是一个有限区间. 由于第一个总体的标准差比较小, 所以它的密度在区间的两端下降得比第二个总体要快.

在两个总体聚集在不同点周围且有不同分散程度的情况下, 在进行分类时, 考虑在中心周围以及中心之间 (即使离中心不是很远, 密度也非常小) 用多元正态分布近似这两个总体是合情理的. 在这种情况下, 我们希望利用一些简单的曲线或曲面把样本空间分成两个分类区域, 最简单的就是使用直线或超平面, 这种分类程式就称为线性方法.

假设 $\boldsymbol{b}$ $(\neq\boldsymbol{0})$ 是一个 (p 元) 向量, c 是一个标量. 当 $\boldsymbol{b}'\boldsymbol{x}\geqslant c$ 时我们把观测 $\boldsymbol{x}$ 归入第一个总体, 当 $\boldsymbol{b}'\boldsymbol{x}<c$ 时归入第二个总体. 我们主要关注这两个总体最大的不同点在于中心不同时的情况, 假设 $\boldsymbol{\mu}^{(1)}\neq\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_2$ 且 $\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2$ 是非奇异的.

当样本来自第 i 个总体时, $\boldsymbol{b}'\boldsymbol{x}$ 服从均值为 $\mathrm{E}(\boldsymbol{b}'\boldsymbol{x}|\pi_i)=\boldsymbol{b}'\boldsymbol{\mu}^{(i)}$、方差为

$$\mathrm{E}(\boldsymbol{b}'\boldsymbol{x}-\boldsymbol{b}'\boldsymbol{\mu}^{(i)})^2|\pi_i=\mathrm{E}[\boldsymbol{b}'(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})(\boldsymbol{x}-\boldsymbol{\mu}^{(i)})'\boldsymbol{b}|\pi_i]=\boldsymbol{b}'\boldsymbol{\Sigma}_i\boldsymbol{b} \tag{6}$$

的一元正态分布. 观测来自第一个总体的误判概率为

$$\begin{aligned}P(2|1)&=\Pr\{\boldsymbol{b}'\boldsymbol{x}<c|\pi_1\}=\Pr\left\{\frac{\boldsymbol{b}'\boldsymbol{x}-\boldsymbol{b}'\boldsymbol{\mu}^{(1)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}}<\frac{c-\boldsymbol{b}'\boldsymbol{\mu}^{(1)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}}\Big|\pi_1\right\}\\&=\Phi\left(\frac{c-\boldsymbol{b}'\boldsymbol{\mu}^{(1)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}}\right)=1-\Phi\left(\frac{\boldsymbol{b}'\boldsymbol{\mu}^{(1)}-c}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}}\right).\end{aligned} \tag{7}$$

观测来自第二个总体的误判概率为

$$\begin{aligned}P(1|2)&=\Pr\{\boldsymbol{b}'\boldsymbol{x}\geqslant c|\pi_2\}=\Pr\left\{\frac{\boldsymbol{b}'\boldsymbol{x}-\boldsymbol{b}'\boldsymbol{\mu}^{(2)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}}\geqslant\frac{c-\boldsymbol{b}'\boldsymbol{\mu}^{(2)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}}\Big|\pi_2\right\}\\&=1-\Phi\left(\frac{c-\boldsymbol{b}'\boldsymbol{\mu}^{(2)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}}\right).\end{aligned} \tag{8}$$

上述概率较小时, 即

$$y_1=\frac{\boldsymbol{b}'\boldsymbol{\mu}^{(1)}-c}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}},\qquad y_2=\frac{c-\boldsymbol{b}'\boldsymbol{\mu}^{(2)}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}} \tag{9}$$

较大时比较理想. 下面我们来考虑在给定 y_2 的条件下使 y_1 变大.

消掉 (9) 式中的 c 得到

$$y_1=[\boldsymbol{b}'\boldsymbol{\gamma}-y_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}](\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{-\frac{1}{2}}, \tag{10}$$

其中 $\boldsymbol{\gamma}=\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)}$. 为了极大化给定 y_2 下的 y_1, 对 y_1 关于 $\boldsymbol{b}$ 求导可得

$$\begin{aligned}\frac{\partial y_1}{\partial \boldsymbol{b}}&=[\boldsymbol{\gamma}-y_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{-\frac{1}{2}}\boldsymbol{\Sigma}_2\boldsymbol{b}](\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{-\frac{1}{2}}\\&\quad-[\boldsymbol{b}'\boldsymbol{\gamma}-y_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}](\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{-\frac{3}{2}}\boldsymbol{\Sigma}_1\boldsymbol{b}\end{aligned} \tag{11}$$

假设

$$t_1=\frac{\boldsymbol{b}'\boldsymbol{\gamma}-y_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}}{\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b}}, \tag{12}$$

$$t_2=\frac{y_2}{\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b}}}, \tag{13}$$

并令 (11) 式为 $\mathbf{0}$ 可得

$$(t_1\boldsymbol{\Sigma}_1+t_2\boldsymbol{\Sigma}_2)\boldsymbol{b}=\boldsymbol{\gamma}. \tag{14}$$

由 (13), (14) 可得 (12). 若存在一对 t_1, t_2, 以及一个向量 $\boldsymbol{b}$ 满足 (12) 和 (13), 由 (9) 式可得 c,

$$c=y_2\sqrt{\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b}}+\boldsymbol{b}'\boldsymbol{\mu}^{(2)}=t_2\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b}+\boldsymbol{b}'\boldsymbol{\mu}^{(2)}. \tag{15}$$

则由 (9), (12) 和 (13) 可得

$$y_1 = \frac{b'\mu^{(1)} - (t_2 b'\Sigma_2 b + b'\mu^{(2)})}{\sqrt{b'\Sigma_1 b}} = t_1\sqrt{b'\Sigma_1 b}. \tag{16}$$

把 (14) 式看作 $t\,(0 \leqslant t \leqslant 1)$ 的函数. 设 $t_1 = t$ 和 $t_2 = 1-t$, 则 $b = (t_1\Sigma_1 + t_2\Sigma_2)^{-1}\gamma$. 定义 $v_1 = t_1\sqrt{b'\Sigma_1 b}$ 和 $v_2 = t_2\sqrt{b'\Sigma_2 b}$. 对 v_1^2 关于 t 求导, 并利用下面的引理可得

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} t^2\gamma'[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Sigma_1[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\gamma \\
=&2t\gamma'[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Sigma_1[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\gamma \\
&-t^2\gamma'[t\Sigma_1 + (1-t)\Sigma_2]^{-1}(\Sigma_1 - \Sigma_2)[t\Sigma_1 + (1-t)\Sigma_2]^{-1} \\
&\cdot\Sigma_1[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\gamma \\
&-t^2\gamma'[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Sigma_1[t\Sigma_1 + (1-t)\Sigma_2]^{-1} \\
&\cdot(\Sigma_1 - \Sigma_2)[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\gamma \\
=&t\gamma'[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Big\{\Sigma_2[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Sigma_1 \\
&+\Sigma_1[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\Sigma_2\Big\}[t\Sigma_1 + (1-t)\Sigma_2]^{-1}\gamma.
\end{aligned} \tag{17}$$

引理 6.10.1 *如果 Σ_1 和 Σ_2 是正定的且 $t_1 > 0, t_2 > 0$, 那么*

$$\Sigma_2[t_1\Sigma_1 + t_2\Sigma_2]^{-1}\Sigma_1 \tag{18}$$

是正定的.

证明 矩阵 (18) 为

$$\Sigma_2[\Sigma_1(t_1\Sigma_2^{-1} + t_2\Sigma_1^{-1})\Sigma_2]^{-1}\Sigma_1 = (t_1\Sigma_2^{-1} + t_2\Sigma_1^{-1})^{-1}. \quad\blacksquare \tag{19}$$

类似地有 $\mathrm{d}v_2^2/\mathrm{d}t < 0$. 因为 $v_1 \geqslant 0, v_2 \geqslant 0$, 所以 v_1 随着 t 的增大而增大, 当 $t=0$ 时 $v_1 = 0$, 当 $t=1$ 时 $v_1 = \sqrt{\gamma'\Sigma_1^{-1}\gamma}$; v_2 关于 t 递减, 当 t 从 0 变化到 1 时, v_2 由 $\sqrt{\gamma'\Sigma_2^{-1}\gamma}$ 递减到 0. 坐标 v_1 和 v_2 是 t 的连续函数. 对于给定的 y_2, $0 \leqslant y_2 \leqslant \sqrt{\gamma'\Sigma_2^{-1}\gamma}$, 存在 t 使得 $y_2 = v_2 = t_2\sqrt{b'\Sigma_2 b}$, 其中当 $t_1 = t, t_2 = 1-t$ 时 b 满足 (14) 式. 因此对于 y_2, $y_1 = v_2 = t_1\sqrt{b'\Sigma_1 b}$ 极大化 y_1. 类似地, 给定 y_1, $0 \leqslant y_1 \leqslant \sqrt{\gamma'\Sigma_1^{-1}\gamma}$, 存在 t 使得 $y_1 = v_1 = t_1\sqrt{b'\Sigma_1 b}$, 而且当 $t_1 = t, t_2 = 1-t$ 时 b 满足 (14) 式. 此时 $y_2 = v_2 = t_2\sqrt{b'\Sigma_2 b}$ 极大化 y_2. 由 $y_1 \geqslant 0, y_2 \geqslant 0$ 可知误判误差不大于 $\frac{1}{2}$.

下面我们证明, 以这种方式定义的 y_1 和 y_2 的集合对应着容许的线性判别程式. 假设 x_1, x_2 在这个集合内, 并且由 z_1, z_2 定义的判别程式严格优于 x_1, x_2, 即 $x_1 \leqslant z_1, x_2 \leqslant z_2$ 且其中至少有一个取严格不等号. 当 $y_1 = z_1$ 时, 设 y_2^* 是 y_2 在所有的线性判别程式中的极大值, 即 $z_1 = y_1, z_2 \leqslant y_2^*$, 因此 $x_1 \leqslant y_1, x_2 \leqslant y_2^*$. 然而, 因为 $\mathrm{d}y_1/\mathrm{d}y_2 < 0$, 所以只有当 $x_1 = y_1, x_2 = y_2^*$ 时这种情况才有可能. 这与假设 z_1, z_2 严格优于 x_1, x_2 矛盾. 因此 x_1, x_2 对应于容许的线性判别程式.

容许线性判别程式的应用

给定 t_1, t_2, 使得 $t_1\boldsymbol{\Sigma}_1 + t_2\boldsymbol{\Sigma}_2$ 正定, 我们通过解线性方程组 (15) 来求解最优的$\boldsymbol{b}$, 然后通过 (9) 式中的任何一个来计算 c. 通常情况下 t_1, t_2 是未知的, 但是可以通过其他的方法得到一个期望的解. 这里有三种方法.

给定一个误判概率的条件下极小化另外一个误判概率

假设 y_2 给定 (或等价地, 给出当样本来自第二个总体时的误判概率), 我们想极大化 y_1(即极小化当样本来自第一个总体时的误判概率). 假设 $y_2 > 0$(即给定的误判概率小于 $\frac{1}{2}$). 若极大值 $y_1 \geqslant 0$, 我们想去找 $t_2 = 1 - t_1$ 使得 $y_2 = t_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}$, 其中 $\boldsymbol{b} = [t_1\boldsymbol{\Sigma}_1 + t_2\boldsymbol{\Sigma}_2]^{-1}\gamma$. 因为 y_2 是 t_2 的增函数, 我们可以通过试错法得到近似解. 当 $t_2 = 0$ 时 $y_2 = 0$; 当 $t_2 = 1$ 时 $y_2 = (\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}} = (\boldsymbol{b}'\gamma)^{\frac{1}{2}} = (\gamma'\boldsymbol{\Sigma}_2^{-1}\gamma)$, 其中 $\boldsymbol{\Sigma}_2\boldsymbol{b} = \gamma$. 我们通过解 (14) 以及插入 $\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b}$, 不断地试用其余的 t_2 值, 直到 $t_2(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}$ 非常接近于期望的值 y_2. [若给定的 $y_2 < (\gamma'\boldsymbol{\Sigma}_2^{-1}\gamma)^{\frac{1}{2}}$, 则 $y_1 > 0$.]

极小化极大判别程式

$y_1 = y_2$ 时的极小化极大判别程式是容许程式. 因为这种程式的两类正确分类的概率都大于 $\frac{1}{2}$, 我们有 $y_1 = y_2 > 0$ 以及 $t_1 > 0, t_2 > 0$. 我们想找 t $(= t_1 = 1 - t_2)$ 满足

$$\begin{aligned} 0 &= y_1^2 - y_2^2 = t^2\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b} - (1-t)^2\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b} \\ &= \boldsymbol{b}'[t^2\boldsymbol{\Sigma}_1 - (1-t)^2\boldsymbol{\Sigma}_2]\boldsymbol{b}. \end{aligned} \tag{20}$$

因为当 t 增加时. y_1^2 随着 t 递增而 y_2^2 随着 t 递减, 所以 (20) 存在唯一解. 这个解可以通过猜测 t $(0 < t < 1)$ 值的试错法来近似. 通过 (14) 式解得 $\boldsymbol{b}$, 再计算 (20) 式右端的二次型, 然后再尝试另外的 t 值.

另一种方法是在 (9) 式中令 $y_1 = y_2$, 然后解 c. $y_1 = y_2$ 的共同值为

$$\frac{\boldsymbol{b}'\boldsymbol{\gamma}}{(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}} + (\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}}, \tag{21}$$

我们想找 $\boldsymbol{b}$ 使得上式极大化, $\boldsymbol{b}$ 具有形式

$$[t\boldsymbol{\Sigma}_1 + (1-t)\boldsymbol{\Sigma}_2]^{-1}\boldsymbol{\gamma}, \tag{22}$$

其中 $0 < t < 1$.

当 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ 时, (21) 式的极大值的二倍为总体间马氏距离的平方. 这意味着当 $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ 时, (21) 式极大值的二倍可以看作是总体间的距离.

Welch and Wimpress(1961) 提出了一个极小化极大判别程式, 并用该程式辨认说话者的声音.

一个先验概率的例子

假设第一个总体和第二个总体的先验概率分别为 q_1 和 q_2. 那么需要极小化的误判概率为

$$q_1[1-\Phi(y_1)]+q_2[1-\Phi(y_2)]=1-[q_1\Phi(y_1)+q_2\Phi(y_2)]. \tag{23}$$

这个解将是一个容许的线性程式. 如果已知 $y_1 \geqslant 0, y_2 \geqslant 0$, 将 $y_1 = t(\boldsymbol{b}'\boldsymbol{\Sigma}_1\boldsymbol{b})^{\frac{1}{2}}, y_2 = (1-t)(\boldsymbol{b}'\boldsymbol{\Sigma}_2\boldsymbol{b})^{\frac{1}{2}}$ 代入 (23) 式, 其中 $\boldsymbol{b} = [t\boldsymbol{\Sigma}_1 + (1-t)\boldsymbol{\Sigma}_2]^{-1}\gamma$. 对 (23) 式关于 t 求导并令其为 0, 可得

$$q_1\phi(y_1)\frac{\mathrm{d}y_1}{\mathrm{d}t}+q_2\phi(y_2)\frac{\mathrm{d}y_2}{\mathrm{d}t}=0, \tag{24}$$

其中 $\phi(u)=(2\pi)^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}u^2}$. 看起来并没有简单或直接的方法解出 (24) 式的 t. (24) 式的左端并不一定是单调的. 事实上, (24) 式可能有许多根. 若是这样, 通过把解代入 (23), 就可以得到绝对极小值. (注意容许的误差概率曲线不一定是凸的).

Anderson and Bahadur(1962) 研究了一般的线性程式, 包括 $y_1 < 0, y_2 < 0$ 的情况. Clunies-Ross and Riffenburgh(1960) 则从几何的角度出发研究了这个问题.

习　　题

6.1 (6.3 节) 设 $\pi_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), i=1,2$. 求容许分类程式的表达式.

6.2 (6.3 节) 证明: 每一个完全程式类都包含容许程式类.

6.3 (6.3 节) 证明: 若容许程式类是完全的, 则必是最小完全类.

6.4 (6.3 节) 由 Neyman-Pearson 基本引理知, 在给定显著性水平的条件下, 原假设$\boldsymbol{x}$来自 $p_1(\boldsymbol{x})$, 备择假设 $\boldsymbol{x}$ 来自 $p_2(\boldsymbol{x})$ 的最优检验具有临界区域 $p_1(\boldsymbol{x})/p_2(\boldsymbol{x}) < k$. 说明 6.3 节的讨论证明了这个结果.

6.5 (6.3 节) 当 $p(\boldsymbol{x}) = n(\boldsymbol{x}|, \boldsymbol{\Sigma})$ 时, 求假设 $\mu = 0$ 对 $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ 的显著性水平为 ε 的最佳检验. 并证明这个检验是对应于备择假设 $\boldsymbol{\mu} = c\boldsymbol{\mu}^*(c>0)$ 的一致最优检验. 最后, 证明不存在对应于备择假设 $\boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}$ 和 $\boldsymbol{\mu} = \boldsymbol{\mu}^{(2)}$ 的一致最优检验, 除非对于某个 $c > 0$ 有 $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$.

6.6 (6.4 节) 设 $P(2|1)$ 和 $P(1|2)$ 分别由 (14) 式和 (15) 式定义. 证明如果 $-\frac{1}{2}\boldsymbol{\Delta}^2 < c < \frac{1}{2}\boldsymbol{\Delta}^2$, 那么 $P(2|1)$ 和 $P(1|2)$ 都是 $\boldsymbol{\Delta}$ 的递减函数.

6.7 (6.4 节) 设 $\boldsymbol{x}' = (\boldsymbol{x}^{(1)'}, \boldsymbol{x}^{(2)'})$. 利用定习题 5.23 和习题 6.6, 证明基于$\boldsymbol{x}$的分类程式组成的类一致优于基于 $\boldsymbol{x}^{(1)}$ 的分类程式组成的类.

6.8 (6.5.1 节) 根据 5.3.4 节给出的数据制定一个准则, 把鸢尾花分类为山鸢尾或变色鸢尾. 对表 3.4 中的 5 个维吉尼亚鸢尾样本进行分类.

6.9 (6.5.1 节) 设 $W(\boldsymbol{x})$ 是由 (2) 式给出的分类准则. 证明: 检验 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}) = N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma})$ 的 T^2 准则与 $W(\bar{\boldsymbol{x}}^{(1)})$ 和 $W(\bar{\boldsymbol{x}}^{(2)})$ 成比例.

6.10 (6.5.1 节) 证明: 随着 N 的增大, $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ (全部都假设来自于 π_1 或者 π_2) 的误判概率减小.

6.11 (6.5 节) 证明: 在变换 (34) 下 $\boldsymbol{M}$ 的元素保持不变, 而且如果充分统计量的任意函数对变换保持不变, 则它必是 $\boldsymbol{M}$ 的函数.

6.12 (6.5 节) 考虑 $\boldsymbol{d}'\boldsymbol{x}^{(i)}$. 求下列比值,

$$\frac{(\mathrm{d}'\bar{\boldsymbol{x}}^{(1)}-\mathrm{d}'\bar{\boldsymbol{x}}^{(2)})^2}{\sum_{\alpha=1}^{N_1}(\mathrm{d}'\boldsymbol{x}_\alpha^{(1)}-\mathrm{d}'\bar{\boldsymbol{x}}^{(1)})^2+\sum_{\alpha=1}^{N_2}(\mathrm{d}'\boldsymbol{x}_\alpha^{(2)}-\mathrm{d}'\bar{\boldsymbol{x}}^{(2)})^2}.$$

6.13 (6.6 节) 证明: (2) 式的导数展开到 n^{-1} 阶项为

$$-\phi\left(\frac{1}{2}\boldsymbol{\Delta}\right)\left\{\frac{1}{2}+\frac{1}{n}\left[\frac{p-1}{\boldsymbol{\Delta}^2}+\frac{p-2}{4}-\frac{p}{8}\boldsymbol{\Delta}^2\right]\right\}.$$

6.14 (6.6 节) 证明: $\mathrm{E}(D^2)$ 是 (4) 式. [提示: 假设 $\boldsymbol{\Sigma}=\boldsymbol{I}$, 证明 $\mathrm{E}(\boldsymbol{S}^{-1}|\boldsymbol{\Sigma}=\boldsymbol{I})=[n/(n-p-1)\boldsymbol{I}]$.]

6.15 (6.6.2 节) 证明:

$$\begin{aligned}&\Pr\left\{\frac{\boldsymbol{Z}-\frac{1}{2}\boldsymbol{D}^2}{\boldsymbol{D}}\leqslant u\middle|\pi_1\right\}-\Pr\left\{\frac{\boldsymbol{Z}-\frac{1}{2}\boldsymbol{\Delta}^2}{\boldsymbol{\Delta}}\leqslant u\middle|\pi_1\right\}\\=&\phi(u)\left\{\frac{1}{2N_1\boldsymbol{\Delta}^2}[u^3+(p-3)u-\boldsymbol{\Delta}^2u+p\boldsymbol{\Delta}]\right.\\&+\frac{1}{2N_2\boldsymbol{\Delta}^2}[u^3+2\boldsymbol{\Delta}u^2+(p-3+\boldsymbol{\Delta}^2)u-\boldsymbol{\Delta}^3+p\boldsymbol{\Delta}]\\&\left.+\frac{1}{4n}[3u^3+4\boldsymbol{\Delta}u^2+(2p-3+\boldsymbol{\Delta}^2)u+2(p-1)\boldsymbol{\Delta}]\right\}+O(n^{-2}).\end{aligned}$$

6.16 (6.8 节) 设 $\pi_i\sim N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})$, $i=1,\cdots,m$. 若 $\boldsymbol{\mu}^{(i)}$ 都在一条直线上 (即 $\boldsymbol{\mu}^{(i)}=\boldsymbol{\mu}+\boldsymbol{\nu}_i\boldsymbol{\beta}$), 证明: 对容许判别程式而言, R_i 由平行平面定义. 并说明这其中只用到一个判别函数 $u_{fk}(\boldsymbol{x})$.

6.17 (6.8 节) 在 8.8 节给出了四个总体的头骨样本数据. 考虑前两个测量和前三个样本. 构造分类函数 $u_{ij}(\boldsymbol{x})$. 求 $q_i=N_i/(N_1+N_2+N_3)$ 时的判别程式. 并求极小化极大判别程式.

6.18 (6.10 节) 证明: 平面方程 $\boldsymbol{b}'\boldsymbol{x}=c$ 在同一点与密度为 π_1 的常量椭球和密度为 π_2 的常量椭球相切.

6.19 (6.8 节) 设 $\boldsymbol{x}_1^{(i)},\cdots,\boldsymbol{x}_{N_i}^{(i)}$ 是来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})(i=1,2,3)$ 的观测, $\boldsymbol{x}$ 是需要被分类的观测. 详细地给出极大似然准则.

6.20 (6.5 节) 证明 (33) 式.

第 7 章　样本协方差阵和样本广义方差的分布

7.1　引　　言

样本协方差阵 $\boldsymbol{S}=[1/(N-1)]\sum_{\alpha=1}^{n}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$ 是总体协方差阵 $\boldsymbol{\Sigma}$ 的无偏估计. 4.2 节给出了矩阵 $\boldsymbol{A}=(N-1)\boldsymbol{S}$ 为 2×2 时的密度. 我们在 7.2 节把这个结果推广到 $\boldsymbol{A}$ 为任意阶矩阵的情况. 当 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 时, 这个分布实际上就是 χ^2 分布的一个推广. $\boldsymbol{A}$ (或者 $\boldsymbol{S}$) 的分布 (经常被称为 Wishart 分布) 是多元统计分析的基础. 在 7.3 节和 7.4 节我们讨论了 Wishart 分布的一些性质.

广义样本方差是样本分散程度的一个度量, 7.5 节将其定义为 $|\boldsymbol{S}|$ 并刻画了它的分布. 7.6 节介绍了观测向量的分量相互独立时的相关系数集的密度.

7.7 节介绍了逆 Wishart 分布, 并将其作为 $\boldsymbol{\Sigma}$ 的先验分布, 从而得到了协方差阵的贝叶斯估计. 7.8 节基于两个损失函数对 $\boldsymbol{\Sigma}$ 的估计 $\boldsymbol{S}$ 进行了改进. 7.9 节讨论了来自椭球等高分布的样本的分布.

7.2　Wishart 分布

下面介绍 $\boldsymbol{A}=\sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}})(\boldsymbol{X}_\alpha-\bar{\boldsymbol{X}})'$ 的分布, 其中 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N(N>p)$ 相互独立, 都服从分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. 如 3.3 节所示, $\boldsymbol{A}$ 与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 分布相同, 其中 $n=N-1$, $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 相互独立而且都服从分布 $N(\boldsymbol{0},\boldsymbol{\Sigma})$. 当 $\boldsymbol{A}$ 正定时, 其密度为

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}\exp(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A})}{2^{\frac{1}{2}np}\pi^{p(p-1)/4}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}. \tag{1}$$

首先考虑 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 的情况. 设

$$(\,\boldsymbol{Z}_1,\quad\cdots,\quad\boldsymbol{Z}_n\,)=\begin{pmatrix}\boldsymbol{v}_1'\\ \vdots\\ \boldsymbol{v}_p'\end{pmatrix}, \tag{2}$$

则 $\boldsymbol{A}=(a_{ij})$ 的元素为这些 n 维向量的内积, $a_{ij}=\boldsymbol{v}_i'\boldsymbol{v}_j$. 向量 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ 相互独立, 都服从分布 $N(\boldsymbol{0},\boldsymbol{I}_n)$. 为方便起见, 我们按照 Gram-Schmidt 正交化方法变换坐标. 设 $\boldsymbol{w}_1=\boldsymbol{v}_1$,

$$\boldsymbol{w}_i=\boldsymbol{v}_i-\sum_{j=1}^{i-1}\boldsymbol{w}_j\frac{\boldsymbol{w}_j'\boldsymbol{v}_i}{\boldsymbol{v}_j'\boldsymbol{w}_j},\quad i=2,\cdots,p. \tag{3}$$

我们归纳证明 $\boldsymbol{w}_k$ 与 $\boldsymbol{w}_i\,(k<i)$ 正交. 假设 $\boldsymbol{w}_k'\boldsymbol{w}_h=0, k\neq h, k,h=1,\cdots,i-1$. (3) 式与 $\boldsymbol{w}_k$ 取内积可得 $\boldsymbol{w}_k'\boldsymbol{w}_i=0, k=1,\cdots,i-1$. (注意到 $\Pr\{\|\boldsymbol{w}_i\|=0\}=0$.)

定义 $t_{ii}=\|\boldsymbol{w}_i\|=\sqrt{\boldsymbol{w}_i'\boldsymbol{w}_i},\ i=1,\cdots,p$, 以及 $t_{ij}=\boldsymbol{v}_i'\boldsymbol{w}_j/\|\boldsymbol{w}_j\|,\ j=1,\cdots,i-1, i=2\cdots,p$. 因为 $\boldsymbol{v}_i=\sum_{j=1}^{i}(t_{ij}/\|\boldsymbol{w}_j\|)\boldsymbol{w}_j$, 所以

$$a_{hi}=\boldsymbol{v}_h'\boldsymbol{v}_i=\sum_{j=1}^{\min(h,i)}t_{hj}t_{ij}. \tag{4}$$

如果我们定义下三角矩阵为 $\boldsymbol{T}=(t_{ij})$, 其中 $t_{ii}>0,\ i=1,\cdots,p, t_{ij}=0,\ i<j$, 则

$$\boldsymbol{A}=\boldsymbol{T}\boldsymbol{T}'. \tag{5}$$

注意到 $t_{ij}\,(j=1,\cdots,i-1)$ 是 $\boldsymbol{v}_i$ 的前 $i-1$ 个坐标, 其中 v_i 在由 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$ 作为前 $i-1$ 个坐标轴所组成的坐标系中. (见图 7.1.) 其他 $n-i+1$ 个坐标的平方和为 $\|\boldsymbol{v}_i\|^2-\sum_{j=1}^{i-1}t_{ij}^2=t_{ii}^2=\|\boldsymbol{w}_i\|^2$. $\boldsymbol{w}_i$ 是 $\boldsymbol{v}_i$ 在 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$ (等价地, 在 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$) 上的投影向量.

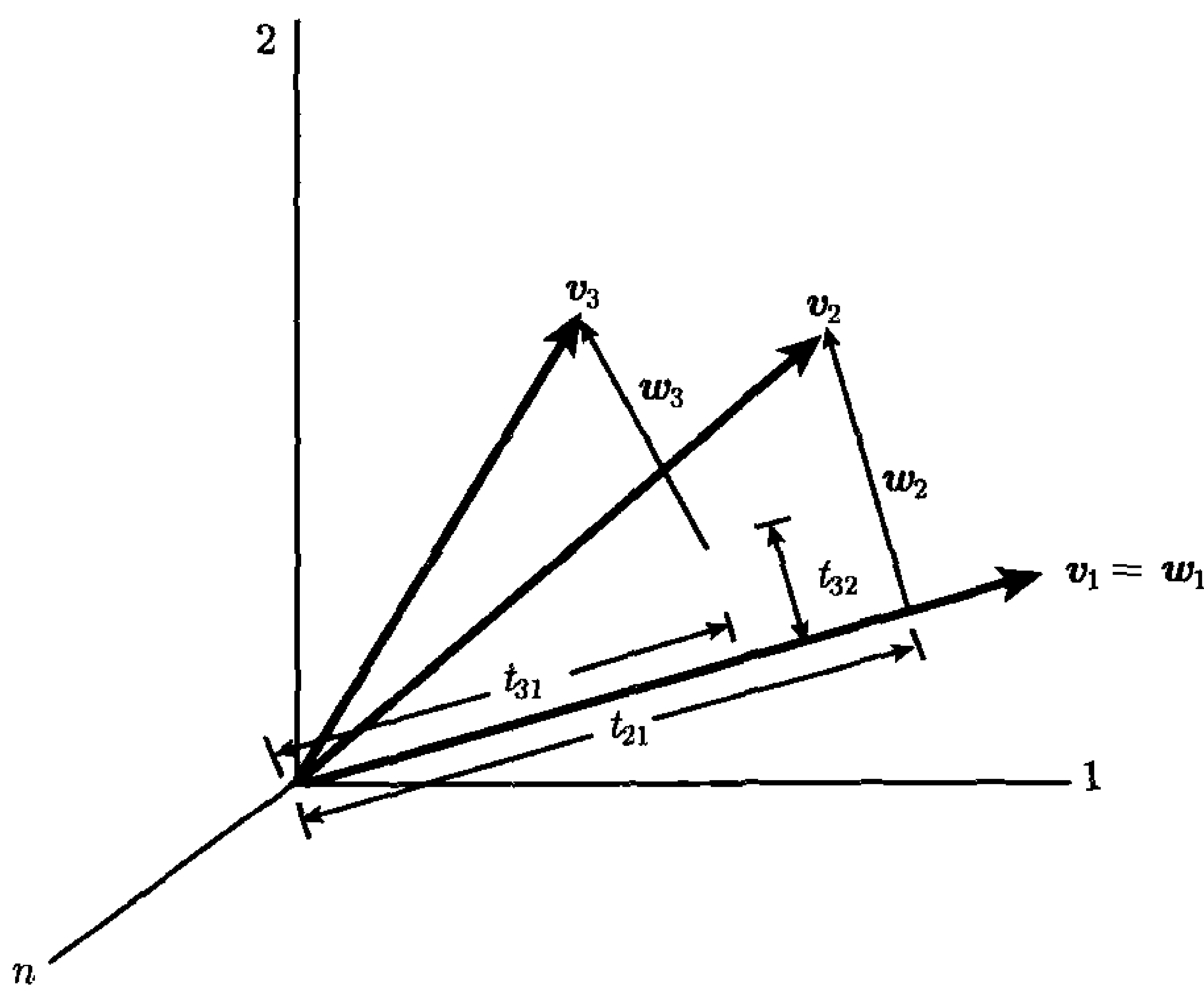


图 7.1 坐标变换

引理 7.2.1 基于 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$ (或等价地, $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$), $t_{i1},\cdots,t_{i,i-1}$ 与 t_{ii}^2 的分布相互独立; $t_{ij}\,(i>j)$ 服从分布 $N(0,1)$; t_{ii}^2 服从自由度为 $n-i+1$ 的 χ^2 分布.

证明 关于以 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$ 定义前 $i-1$ 个坐标轴的新正交坐标, $\boldsymbol{v}_i$ 的坐标是相互独立的均值为 0 方差为 1 的正态分布 (定理 3.3.1). t_{ii}^2 是去掉前 $i-1$ 个坐标的坐标平方和. ■

推论 7.2.1 设 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n\,(n\geqslant p)$ 独立且都服从分布 $N(\boldsymbol{0},\boldsymbol{I})$; 设 $\boldsymbol{A}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha=\boldsymbol{T}\boldsymbol{T}'$, 其中 $t_{ij}=0\,(i<j)$ 和 $t_{ii}>0\,(i=1,\cdots,p)$. 则 $t_{11},t_{21},\cdots,t_{pp}$

是独立分布的; $t_{ij}\,(i>j)$ 服从 $N(0,1)$ 分布; t_{ii}^2 服从自由度为 $n-i+1$ 的 χ^2 分布.

因为 t_{ii} 的密度为 $2^{-\frac{1}{2}(n-i-1)}t^{n-i}\mathrm{e}^{-\frac{1}{2}t^2}/\Gamma[\frac{1}{2}(n+1-i)]$, $t_{ij}\,(j=1,\cdots,i,\ i=1,\cdots,p)$ 的联合密度为

$$\prod_{i=1}^{p}\frac{t_{ii}^{n-i}\exp(-\frac{1}{2}\sum_{j=1}^{i}t_{ij}^2)}{\pi^{\frac{1}{2}(i-1)}2^{\frac{1}{2}n-1}\Gamma[\frac{1}{2}(n+1-i)]} \tag{6}$$

$$=\frac{\prod_{i=1}^{p}t_{ii}^{n-i}\exp(-\frac{1}{2}\prod_{i=1}^{p}\sum_{j=1}^{i}t_{ij}^2)}{2^{\frac{1}{2}p(n-2)}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}.$$

设 $\boldsymbol{C}$ 为下三角矩阵 $(c_{ij}=0,\ i<j)$ 且满足 $\boldsymbol{\Sigma}=\boldsymbol{C}\boldsymbol{C}'$ 和 $c_{ii}>0$. 线性变换 $\boldsymbol{T}^*=\boldsymbol{C}\boldsymbol{T}$, 即

$$t_{ij}^*=\sum_{k=j}^{i}c_{ik}t_{kj},\quad i\geqslant j, \tag{7}$$

$$=0,\quad i<j,$$

可以写为

$$\begin{pmatrix} t_{11}^* \\ t_{21}^* \\ t_{22}^* \\ \vdots \\ t_{p1}^* \\ \vdots \\ t_{pp}^* \end{pmatrix}=\begin{pmatrix} c_{11} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ \boldsymbol{x} & c_{22} & 0 & \cdots & 0 & \cdots & 0 \\ \boldsymbol{x} & \boldsymbol{x} & c_{22} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} & \cdots & c_{pp} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \boldsymbol{x} & \boldsymbol{x} & \boldsymbol{x} & \cdots & \boldsymbol{x} & \cdots & c_{pp} \end{pmatrix}\begin{pmatrix} t_{11} \\ t_{21} \\ t_{22} \\ \vdots \\ t_{p1} \\ \vdots \\ t_{pp} \end{pmatrix}, \tag{8}$$

其中 $\boldsymbol{x}$ 表示一个可能非零的元素. 由于变换矩阵为三角矩阵, 所以它的行列式为对角元的乘积, 即 $\prod_{i=1}^{p}c_{ii}^{i}$. 由$\boldsymbol{T}$到 $\boldsymbol{T}^*$ 的变换的雅可比行列式是该行列式的倒数. 把 $t_{ii}=t_{ii}^*/c_{ii}$ 代入 (6), 利用

$$\sum_{i=1}^{p}\sum_{j=1}^{i}t_{ij}^2=\mathrm{tr}\boldsymbol{T}\boldsymbol{T}' \tag{9}$$

$$=\mathrm{tr}\boldsymbol{C}^{-1}\boldsymbol{T}^*\boldsymbol{T}^{*'}(\boldsymbol{C}^{-1})'$$

$$=\mathrm{tr}\boldsymbol{T}^*\boldsymbol{T}^{*'}\boldsymbol{C}^{'-1}\boldsymbol{C}^{-1}$$

$$=\mathrm{tr}\boldsymbol{T}^*\boldsymbol{T}^{*'}\boldsymbol{\Sigma}^{-1}=\mathrm{tr}\boldsymbol{T}^{*'}\boldsymbol{\Sigma}^{-1}\boldsymbol{T}^*$$

和 $\prod_{i=1}^{p}c_{ii}^2=|\boldsymbol{C}||\boldsymbol{C}'|=|\boldsymbol{\Sigma}|$, 就可以得到 $\boldsymbol{T}^*$ 的密度.

定理 7.2.1　设 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n(n>p)$ 独立同分布, 且都服从 $N(0,\boldsymbol{\Sigma})$; 设 $\boldsymbol{A}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'$ 和 $\boldsymbol{A}=\boldsymbol{T}^*\boldsymbol{T}^{*'}$, 其中 $t_{ij}^*=0\,(i<j)$ 和 $t_{ii}^*>0$. 则 $\boldsymbol{T}^*$ 的密度为

$$\frac{\prod_{i=1}^{p}t_{ii}^{*n-i}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{T}^*\boldsymbol{T}^{*'}}}{2^{\frac{1}{2}p(n-2)}\pi^{p(p-1)/4}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\prod_{i=1}^{p}\Gamma[\frac{1}{2}(n+1-i)]}. \tag{10}$$

当 $h \geqslant i$ 时, 我们可以把 (4) 式写为 $a_{hi} = \sum_{j=1}^{i} t_{hj}^* t_{ij}^*$. 这时

$$\frac{\partial a_{hi}}{\partial t_{kl}^*} = 0, \quad k > h, \tag{11}$$
$$= 0, \quad k = h,\ l > i;$$

即, 当 k, l 以字典顺序超过 h, i 时 $\partial a_{hi}/\partial t_{kl}^* = 0$. 由 $\boldsymbol{A}$ 到 $\boldsymbol{T}^*$ 的变换的雅可比行列式是对角元为

$$\frac{\partial a_{hh}}{\partial t_{hh}^*} = 2t_{hh}^*, \tag{12}$$

$$\frac{\partial a_{hi}}{\partial t_{hi}^*} = t_{ii}^*, \quad h > i \tag{13}$$

的下三角矩阵的行列式. 因此雅可比行列式为 $2^p \prod_{i=1}^{p} t_{ii}^{*p+1-i}$. $\boldsymbol{T}^*$ 到 $\boldsymbol{A}$ 的变换的雅可比行列式是其倒数.

定理 7.2.2 设 $\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n$ 独立同分布, 且都服从 $N(0, \boldsymbol{\Sigma})$. 则当 $\boldsymbol{A} = \sum_{\alpha=1}^{n} \boldsymbol{Z}_\alpha \boldsymbol{Z}_\alpha'$ 正定时, 其密度为

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}}}{2^{\frac{1}{2}pn} \pi^{p(p-1)/4} |\boldsymbol{\Sigma}|^{\frac{1}{2}n} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)]}, \tag{14}$$

否则为零.

推论 7.2.2 设 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N\ (N > p)$ 独立同分布, 且都服从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 则 $\boldsymbol{A} = \sum_{\alpha=1}^{N} (\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}})(\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}})'$ 的密度为 (14) 式, 其中 $n = N - 1$.

把 (14) 式的密度记为 $w(\boldsymbol{A}|\boldsymbol{\Sigma}, n)$, 联合密度记为 $W(\boldsymbol{\Sigma}, n)$. 若 $n < p$, $\boldsymbol{A}$是没有密度的, 但是我们仍然把它的密度定义为 $W(\boldsymbol{\Sigma}, n)$.

推论 7.2.3 设 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N\ (N > p)$ 独立同分布, 且都服从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. $\boldsymbol{S} = (1/n) \sum_{\alpha=1}^{N} (\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}})(\boldsymbol{X}_\alpha - \bar{\boldsymbol{X}})'$ 的密度为 $W[(1/n)\boldsymbol{\Sigma}, n]$, 其中 $n = N - 1$.

证明 $\boldsymbol{S}$ 与 $\sum_{\alpha=1}^{n} [1/\sqrt{n} \boldsymbol{Z}_\alpha][1/\sqrt{n} \boldsymbol{Z}_\alpha]'$ 的分布相同, 其中 $(1/\sqrt{n})\boldsymbol{Z}_1, \cdots, (1/\sqrt{n})\boldsymbol{Z}_N$ 相互独立且都服从 $N(0, (1/n)\boldsymbol{\Sigma})$ 分布. 由定理 7.2.2 可得该推论. ■

4.2.1 节中给出的 $p = 2$ 时的 Wishart 分布是 Fisher(1915) 发现的. Wishart(1928) 通过几何方法利用上面定义的 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p$ 得到了任意 p 时的分布. 就像 3.2 节说明的那样, $\boldsymbol{A}$ 的第 i 个对角元是第 i 个向量长度的平方, $a_{ii} = \boldsymbol{v}_i' \boldsymbol{v}_i = \|\boldsymbol{v}_i\|^2$, 第 i, j 个非对角元是向量 $\boldsymbol{v}_i$ 与 $\boldsymbol{v}_j$ 的长度以及它们的夹角余弦的乘积. 矩阵 $\boldsymbol{A}$ 确定了这些向量的长度及结构.

下面给出当 $\boldsymbol{\Sigma} = \boldsymbol{I}$ 时, 直角坐标 $t_{ij}\ (i > j)$ 密度推导的几何解释①. t_{11} 的概率元近似为 $\|\boldsymbol{v}_1\|$ 落入区间 $t_{11} < \|\boldsymbol{v}_1\| < t_{11} + \mathrm{d}t_{11}$ 的概率, 也就是 $\boldsymbol{v}_1$ 落入内半径为 t_{11} 且厚度为 $\mathrm{d}t_{11}$ 的 n 维球壳的概率. 在这个区域, 密度 $(2\pi)^{-\frac{1}{2}n} \exp(-\frac{1}{2}\boldsymbol{v}_1'\boldsymbol{v}_1)$ 近似为常量, 即 $(2\pi)^{-\frac{1}{2}n} \exp(-\frac{1}{2}t_{11}^2)$. n 维单位球的曲面面积为 $C(n) = 2\pi^{\frac{1}{2}n}/\Gamma(\frac{1}{2}n)$

① 在本书的第 1 版中, Wishart 分布的推导以及它的几何解释都是基于非正交向量 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_p$ 的.

(习题 7.1~7.3), 球壳的体积近似为 $C(n)t_{11}^{n-1}\mathrm{d}t_{11}$. 概率元是体积和近似密度的乘积, 即

$$2^{-(\frac{1}{2}n-1)}t_{11}^{n-1}\exp\left(-\frac{1}{2}t_{11}^2\right)\mathrm{d}t_{11}/\Gamma\left(\frac{1}{2}n\right). \tag{15}$$

给定 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ (即给定 $w_1,\cdots,w_{i-1}$) 时, $t_{i1},\cdots,t_{i,i-1},t_{ii}$ 的概率元近似为 $\boldsymbol{v}_i$ 落入区域 $t_{i1}<\boldsymbol{v}_i'\boldsymbol{w}_1/\|\boldsymbol{w}_1\|<t_{i1}+\mathrm{d}t_{i1},\cdots,t_{i,i-1}<\boldsymbol{v}_i'\boldsymbol{w}_{i-1}/\|\boldsymbol{w}_{i-1}\|<t_{i,i-1}+\mathrm{d}t_{i,i-1}$ 和 $t_{ii}<\|\boldsymbol{w}_i\|<t_{ii}-\mathrm{d}t_{ii}$ 的概率, 其中 $\boldsymbol{w}_i$ 是 $\boldsymbol{v}_i$ 到正交于 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$ 的 $n-i+1$ 维空间上的投影. 前 $i-1$ 对不等式的每一对都定义了两个超平面之间的区域 (不同对之间是正交的). 最后一对不等式定义了一个圆柱形的壳, 它与由 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$ 张成的 $i-1$ 维超平面的交集为 $n-i+1$ 维内半径为 t_{ii} 的球壳. 在这个区域中, 密度 $(2\pi)^{-\frac{1}{2}n}\exp(-\frac{1}{2}\boldsymbol{v}_i'\boldsymbol{v}_i)$ 近似为常量 $(2\pi)^{-\frac{1}{2}n}\exp(-\frac{1}{2}\sum_{j=1}^{i}t_{ij}^2)$. 该区域的体积近似为 $\mathrm{d}t_{i1}\cdots\mathrm{d}t_{i,i-1}C(n-i+1)t_{ii}^{n-i}\mathrm{d}t_{ii}$. 概率元为

$$\frac{2^{-(\frac{1}{2}n-1)}\pi^{-\frac{1}{2}(i-1)}t_{ii}^{n-i}\exp(-\frac{1}{2}\sum_{j=1}^{i}t_{ij}^2)}{\Gamma[\frac{1}{2}(n+1-i)]}\mathrm{d}t_{i1}\cdots\mathrm{d}t_{ii}. \tag{16}$$

则 (15) 式和 (16) 式 $i=2,\cdots,p$ 的乘积就是 (6) 式乘以 $\mathrm{d}t_{11}\cdots\mathrm{d}t_{pp}$.

这种解释由 Sverdrup(1947) [$p=3$ 的情况由 Fog(1948) 给出] 给出, 等价于 Wishart[以及后来的 Mahalanobis, Bose, and Roy (1937)] 的几何推导方法. Madow (1938) 提出了另外一种方法, 即利用相关系数 ($\boldsymbol{\Sigma}=\boldsymbol{I}$) 的分布, 该相关系数分布是 Hotelling 通过考虑偏相关系数而得到的. Hsu (1939b) 提供了一个归纳证明, Rasch (1948) 给出了一种涉及函数方程的方法. Ingham (1933) 和 Wishart and Bartlett (1933) 应用其他的方法得到特征函数及其倒数.

Cramér (1946) 验证了 Wishart 分布具有$\boldsymbol{A}$的特征函数. 利用另外的矩阵变换, Elfving (1947), Mauldon (1955) 和 Olkin and Roy (1954) 通过 Bartlett 分解得到了 Wishart 分布. Kshirsagar (1959) 把他的推导建立在随机正交变换上. Narain (1948), (1950) 和 Ogawa (1953) 应用了回归方法. James (1954), Khatri and Ramachandran (1958) 和 Khatri (1963) 应用了不同的方法. Giri (1977) 利用了不变性. Wishart (1948) 对当时的推导方法进行了概括论述. 这其中的一些方法在习题中有所涉及.

称 $\boldsymbol{A}=\boldsymbol{T}\boldsymbol{T}'$ 为 Bartlett 分解 [Bartlett(1939)], Mahatanobis, Bose, and Roy (1937) 把 $\boldsymbol{T}$ 中的 (非零) 元素称为直角坐标.

推论 7.2.4

$$\int\underset{\boldsymbol{B}>0}{\cdots}\int|\boldsymbol{B}|^{t-\frac{1}{2}(p+1)}\mathrm{e}^{-\mathrm{tr}\boldsymbol{B}}\mathrm{d}\boldsymbol{B}=\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left[t-\frac{1}{2}(i-1)\right]. \tag{17}$$

证明 这里 $\boldsymbol{B}>0$ 表示 $\boldsymbol{B}$ 正定. 由于 (14) 式是一个密度, 所以对 $\boldsymbol{A}>0$ 积分, 其值应为 1. 令 $\boldsymbol{\Sigma}=\boldsymbol{I},\boldsymbol{A}=2\boldsymbol{B}$ ($\mathrm{d}\boldsymbol{A}=2\mathrm{d}\boldsymbol{B}$) 和 $n=2t$. 则同理, 对 (17) 式关于

半整数 t 积分, 其值也为 1. 然而, 如果由 (6) 式来推导 (14) 式, 可以令 n 为大于 $p-1$ 的任意实数. 实际上 (17) 式对于任意满足 $\text{Re}(t) > p-1$ ($\text{Re}(t)$ 表示 t 的实数部分) 的复数 t 都成立. ■

定义 7.2.1 *多元伽玛函数定义为*

$$\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left[t - \frac{1}{2}(i-1)\right]. \tag{18}$$

Wishart 密度可以写为

$$w(\boldsymbol{A}|\boldsymbol{\Sigma}, n) = \frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}}}{2^{\frac{1}{2}pn} |\boldsymbol{\Sigma}|^{\frac{1}{2}n} \Gamma_p(\frac{1}{2}n)}. \tag{19}$$

7.3 Wishart 分布的一些性质

7.3.1 特征函数

Wishart 分布的特征函数可以由观测的分布直接得到. 设 $\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n$ 分布独立, 每一个都具有密度

$$\frac{1}{(2\pi)^{\frac{1}{2}p} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} z' \boldsymbol{\Sigma}^{-1} z\right). \tag{1}$$

设

$$\boldsymbol{A} = \sum_{\alpha=1}^{n} \boldsymbol{Z}_\alpha \boldsymbol{Z}'_\alpha. \tag{2}$$

引入 $p \times p$ 的矩阵 $\boldsymbol{\Theta} = (\theta_{ij})$, 其中 $\theta_{ij} = \theta_{ji}$. $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}, \cdots, \boldsymbol{A}_{pp}, 2\boldsymbol{A}_{12}, 2\boldsymbol{A}_{13}, \cdots, 2\boldsymbol{A}_{p-1,p}$ 的特征函数为

$$\begin{aligned} \mathrm{E}\Big[\exp[\mathrm{i}\,\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Theta})]\Big] &= \mathrm{E}\left[\exp\left(\mathrm{i}\,\mathrm{tr}\sum_{\alpha=1}^{n} \boldsymbol{Z}_\alpha \boldsymbol{Z}'_\alpha \boldsymbol{\Theta}\right)\right] \\ &= \mathrm{E}\left[\exp\left(\mathrm{i}\,\mathrm{tr}\sum_{\alpha=1}^{n} \boldsymbol{Z}'_\alpha \boldsymbol{\Theta} \boldsymbol{Z}_\alpha\right)\right] \\ &= \mathrm{E}\left[\exp\left(\mathrm{i}\sum_{\alpha=1}^{n} \boldsymbol{Z}'_\alpha \boldsymbol{\Theta} \boldsymbol{Z}_\alpha\right)\right]. \end{aligned} \tag{3}$$

由引理 2.6.1 可得

$$\mathrm{E}\left[\exp\left(\mathrm{i}\sum_{\alpha=1}^{n} \boldsymbol{Z}'_\alpha \boldsymbol{\Theta} \boldsymbol{Z}_\alpha\right)\right] = \prod_{\alpha=1}^{n} \mathrm{E}\Big[\exp(\mathrm{i}\boldsymbol{Z}'_\alpha \boldsymbol{\Theta} \boldsymbol{Z}_\alpha)\Big] = \Big\{\mathrm{E}\Big[\exp(\mathrm{i}\boldsymbol{Z}'\boldsymbol{\Theta}\boldsymbol{Z})\Big]\Big\}^n, \tag{4}$$

其中 $\boldsymbol{Z}$ 的密度为 (1). 当 $\boldsymbol{\Theta}$ 为实矩阵时, 存在非奇异矩阵 $\boldsymbol{B}$, 满足

$$\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B} = \boldsymbol{I}, \tag{5}$$

$$\boldsymbol{B}'\boldsymbol{\Theta}\boldsymbol{B} = \boldsymbol{D}, \tag{6}$$

其中$\boldsymbol{D}$为实的对角矩阵 (附录中的定理 A.2.2). 令 $z=\boldsymbol{B}y$, 则由引理 2.6.2 知

$$\begin{aligned}\mathrm{E}\Big[\exp(\mathrm{i}\boldsymbol{Z}'\boldsymbol{\Theta}\boldsymbol{Z})\Big] &= \mathrm{E}\Big[\exp(\mathrm{i}\boldsymbol{Y}'\boldsymbol{D}\boldsymbol{Y})\Big] \\ &= \mathrm{E}\Big[\prod_{j=1}^{p}\exp(\mathrm{i}d_{jj}\boldsymbol{Y}_j^2)\Big] \\ &= \prod_{j=1}^{p}\mathrm{E}\Big[\exp(\mathrm{i}d_{jj}\boldsymbol{Y}_j^2)\Big].\end{aligned} \tag{7}$$

在第二个乘积中的第 j 个因子为 $\mathrm{E}[\exp(\mathrm{i}d_{jj}\boldsymbol{Y}_j^2)]$, 其中 $\boldsymbol{Y}_j$ 服从分布 $N(0,1)$; 这是自由度为 1 的 χ^2 分布的特征函数, 即 $(1-2\mathrm{i}d_{jj})^{-\frac{1}{2}}$[通过把 $\exp(\mathrm{i}d_{jj}\boldsymbol{y}_j^2)$ 展开为幂级数, 然后利用逐项积分的方法来证明]. 由于 $\boldsymbol{I}-2\mathrm{i}\boldsymbol{D}$ 为对角矩阵, 则

$$\mathrm{E}\Big[\exp(\mathrm{i}\boldsymbol{Z}'\boldsymbol{\Theta}\boldsymbol{Z})\Big] = \prod_{j=1}^{p}(1-2\mathrm{i}d_{jj})^{-\frac{1}{2}} = |\boldsymbol{I}-2\mathrm{i}\boldsymbol{D}|^{-\frac{1}{2}}. \tag{8}$$

由 (5) 式和 (6) 式可得

$$\begin{aligned}|\boldsymbol{I}-2\mathrm{i}\boldsymbol{D}| &= |\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}-2\mathrm{i}\boldsymbol{B}'\boldsymbol{\Theta}\boldsymbol{B}| \\ &= |\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1}-2\mathrm{i}\boldsymbol{\Theta})\boldsymbol{B}| \\ &= |\boldsymbol{B}'|\cdot|\boldsymbol{\Sigma}^{-1}-2\mathrm{i}\boldsymbol{\Theta}|\cdot|\boldsymbol{B}| \\ &= |\boldsymbol{B}|^2\cdot|\boldsymbol{\Sigma}^{-1}-2\mathrm{i}\boldsymbol{\Theta}|,\end{aligned} \tag{9}$$

$|\boldsymbol{B}'|\cdot|\boldsymbol{\Sigma}^{-1}|\cdot|\boldsymbol{B}|=|\boldsymbol{I}|=1$, $|\boldsymbol{B}|^2=1/|\boldsymbol{\Sigma}^{-1}|$. 联立上面的结果可得

$$\mathrm{E}\Big[\exp(\mathrm{i}\,\mathrm{tr}\boldsymbol{A}\boldsymbol{\Theta})\Big] = \frac{|\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}n}}{|\boldsymbol{\Sigma}^{-1}-2\mathrm{i}\boldsymbol{\Theta}|^{\frac{1}{2}n}} = |\boldsymbol{I}-2\mathrm{i}\boldsymbol{\Theta}\boldsymbol{\Sigma}|^{-\frac{1}{2}n}. \tag{10}$$

可证这个结果在 $(\mathrm{Re}(\sigma^{ik}-2\mathrm{i}\theta_{jk}))$ 正定时成立, 尤其对于所有实值 $\boldsymbol{\Theta}$ 都是成立的. 当 $\boldsymbol{\Sigma}$ 为奇异矩阵时, 这个结果也是成立的.

定理 7.3.1　*若 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 相互独立, 且都服从分布 $N(0,\boldsymbol{\Sigma})$, 则 $\boldsymbol{A}_{11},\boldsymbol{A}_{22},\cdots,\boldsymbol{A}_{pp},2\boldsymbol{A}_{12},2\boldsymbol{A}_{13},\cdots,2\boldsymbol{A}_{p-1,p}$ 的特征函数为 (10) 式, 其中 $(A_{ij})=\boldsymbol{A}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$.*

7.3.2　Wishart 矩阵的和

设 $\boldsymbol{A}_i\,(i=1,2)$ 分布独立且分别服从 $W(\boldsymbol{\Sigma},n_i)$. 则 $\boldsymbol{A}_1$ 的分布与 $\sum_{\alpha=1}^{n_1}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 的分布相同, $\boldsymbol{A}_2$ 的分布与 $\sum_{\alpha=n_1+1}^{n_1+n_2}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 的分布相同, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{n_1+n_2}$ 相互独立而且都服从分布 $N(0,\boldsymbol{\Sigma})$. 则 $\boldsymbol{A}=\boldsymbol{A}_1+\boldsymbol{A}_2$ 的分布与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 的分布相同, 其中 $n=n_1+n_2$. 因此 $\boldsymbol{A}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布. 同理, q 个分布独立且协方差都为 $\boldsymbol{\Sigma}$ 的 Wishart 矩阵之和仍服从 Wishart 分布, 且其协方差为 $\boldsymbol{\Sigma}$, 自由度为这 q 个 Wishart 矩阵自由度的和.

定理 7.3.2 若 $A_1, \cdots, A_q$ 相互独立且 A_i 分别服从 $W(\Sigma, n_i)$ 分布, 则

$$A = \sum_{i=1}^{q} A_i \tag{11}$$

的分布为 $W(\Sigma, \sum_{i=1}^{q} n_i)$.

7.3.3 一种线性变换

我们会经常用到变换

$$A = CBC', \tag{12}$$

其中C为非奇异的 $p \times p$ 矩阵. 若A服从 $W(\Sigma, n)$ 分布, 则B的分布为 $W(\Phi, n)$, 其中

$$\Phi = C^{-1}\Sigma C'^{-1}, \tag{13}$$

该结论证明如下: 设 $A = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, 其中 $Z_1, \cdots, Z_n$ 相互独立且都服从 $N(0, \Sigma)$ 分布. 则 $Y_\alpha = C^{-1}Z_\alpha$ 服从 $N(0, \Phi)$. 然而

$$B = \sum_{\alpha=1}^{n} Y_\alpha Y_\alpha' = C^{-1}\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha' C'^{-1} = C^{-1}AC'^{-1} \tag{14}$$

的分布为 $W(\Phi, n)$. 最后, 变换 (12) 的雅可比行列式 $|\partial(A)/\partial(B)|$ 为

$$\left|\frac{\partial(A)}{\partial(B)}\right| = \frac{w(B, \Phi, n)}{w(A, \Sigma, n)} = \frac{|B|^{\frac{1}{2}(n-p-1)}|\Sigma|^{\frac{1}{2}n}}{|A|^{\frac{1}{2}(n-p-1)}|\Phi|^{\frac{1}{2}n}} = \mathrm{mod}|C|^{p+1}. \tag{15}$$

定理 7.3.3 由 A 到 B 的变换 (12) 的雅可比行列式为 $|C|^{p+1}$, 其中 A 和 B 为对称矩阵.

7.3.4 边际分布

若 A 服从分布 $W(\Sigma, n)$, 则求得 A 的任意子集的边际分布是困难的. 尽管如此, 一些子集的边际分布还是容易得到的. 下面的两个定理给出了一些边际分布.

定理 7.3.4 设 A 和 Σ 分别被分为 $q, p-q$ 行和列

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{16}$$

若 A 服从 $W(\Sigma, n)$ 分布, 则 A_{11} 的分布为 $W(\Sigma_{11}, n)$.

证明 A 的分布与 $\sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$ 的分布相同, 其中 $Z_1, \cdots, Z_n$ 分布独立且都服从 $N(0, \Sigma)$. 把 Z_α 分成两部分, 每部分分别有 q 和 $p-q$ 个分量, $Z_\alpha = (Z_\alpha^{(1)\prime}, Z_\alpha^{(2)\prime})'$. 则 $Z_1^{(1)}, \cdots, Z_n^{(1)}$ 相互独立且都服从 $N(0, \Sigma_{11})$ 分布, 从而 A_{11} 的分布与 $\sum_{\alpha=1}^{n} Z_\alpha^{(1)} Z_\alpha^{(1)\prime}$ 的分布相同, 即有分布 $W(\Sigma_{11}, n)$. ∎

定理 7.3.5 设 A 和 Σ 被分块为 $p_1, \cdots, p_q$ 行和列 $(p_1 + \cdots + p_q = p)$,

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1q} \\ \vdots & & \vdots \\ A_{q1} & \cdots & A_{qq} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \cdots & \Sigma_{1q} \\ \vdots & & \vdots \\ \Sigma_{q1} & \cdots & \Sigma_{qq} \end{pmatrix}. \tag{17}$$

如果 $\boldsymbol{A}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布且 $\boldsymbol{\Sigma}_{ij}=\mathbf{0}$, $i\neq j$, 则 $\boldsymbol{A}_{11},\boldsymbol{A}_{22},\cdots,\boldsymbol{A}_{qq}$ 是分布独立的, 且 $\boldsymbol{A}_{jj}$ 服从 $W(\boldsymbol{\Sigma}_{jj},n)$ 分布.

证明　$\boldsymbol{A}$ 的分布与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$ 的分布相同, 其中 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 分布独立且都服从 $N(0,\boldsymbol{\Sigma})$. 把 $\boldsymbol{Z}_\alpha$ 做类似于 $\boldsymbol{A}$ 和$\boldsymbol{\Sigma}$的分块,

$$\boldsymbol{Z}_\alpha=\begin{pmatrix}\boldsymbol{Z}_\alpha^{(1)}\\ \vdots\\ \boldsymbol{Z}_\alpha^{(q)}\end{pmatrix}. \tag{18}$$

因为 $\boldsymbol{\Sigma}_{ij}=0$, 所以 $\boldsymbol{Z}_1^{(1)},\cdots,\boldsymbol{Z}_n^{(1)},\cdots,\boldsymbol{Z}_1^{(q)},\cdots,\boldsymbol{Z}_n^{(q)}$ 是相互独立的. 则 $\boldsymbol{A}_{11}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha^{(1)}\boldsymbol{Z}_\alpha^{(1)'}$, $\ldots$,$\boldsymbol{A}_{qq}=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha^{(q)}\boldsymbol{Z}_\alpha^{(q)'}$ 是相互独立的. 从而由定理 7.3.4 可得定理 7.3.5. ■

7.3.5　条件分布

在 4.3 节, 我们讨论了在给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时, $\boldsymbol{X}^{(1)}$ 的条件分布的参数估计问题. 将定理 7.2.2 应用到定理 4.3.3 得到如下定理.

定理 7.3.6　类似于 (16) 式, 把 $\boldsymbol{A}$ 和 $\boldsymbol{\Sigma}$ 分为 $q,p-q$ 行和列, 如果 $\boldsymbol{A}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布, 则 $\boldsymbol{A}_{11\cdot2}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ 的分布是 $W(\boldsymbol{\Sigma}_{11\cdot2},n-p+q)$, $n\geqslant p-q$.

注意到, 定理 7.3.6 隐含着, 无论 $\boldsymbol{\Sigma}$ 如何, $\boldsymbol{A}_{11\cdot2}$ 与 $\boldsymbol{A}_{22}$ 和 $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$ 都是独立分布的.

7.4　Cochran 定理

Cochran 定理 [Cochran(1934)] 在证明某一向量的二次型与向量的平方和具有相同分布时特别有用. 它是一个代数定理的统计论述, 这个代数定理就是下面我们要给出的引理.

引理 7.4.1　若 $N\times N$ 对称矩阵 $\boldsymbol{C}_i$ 的秩为 r_i, $i=1,\cdots,m$, 且

$$\sum_{i=1}^{m}\boldsymbol{C}_i=\boldsymbol{I}_N, \tag{1}$$

则

$$\sum_{i=1}^{m}r_i=N \tag{2}$$

的一个充分必要条件为: 存在一个 $N\times N$ 正交矩阵 $\boldsymbol{P}$ 使得对于 $i=1,\cdots,m$, 有

$$\boldsymbol{P}\boldsymbol{C}_i\boldsymbol{P}'=\begin{pmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\boldsymbol{I}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}\end{pmatrix}, \tag{3}$$

其中 $\boldsymbol{I}$ 为 r_i 阶, 左上角的 $\mathbf{0}$ 为 $\sum_{j=1}^{i-1}r_j$ (当 $i=1$ 时, 此项不存在) 阶方阵, 右下角的 $\mathbf{0}$ 为 $\sum_{j=i+1}^{m}r_j$ (当 $i=m$ 时, 此项不存在) 阶方阵.

证明 由 (1) 式知, 对于任意 $i=1,\cdots,m$, (3) 式的和为 $\boldsymbol{I}_N$, 从而可得必要性. 下面我们证明充分性; 假设 (2) 式成立. 存在正交矩阵 $\boldsymbol{P}_j$, 使得 $\boldsymbol{P}_j\boldsymbol{C}_j\boldsymbol{P}_j'$ 为对角矩阵, 并且对角元为 $\boldsymbol{C}_j$ 的特征根. 非零特征根的个数为 $\boldsymbol{C}_j$ 的秩, 即 r_i 个, 零特征根的个数为 $N-r_i$ 个. 把 $\boldsymbol{P}_j\boldsymbol{C}_j\boldsymbol{P}_j'$ 表示为

$$\boldsymbol{P}_j\boldsymbol{C}_j\boldsymbol{P}_j'=\begin{pmatrix}0&0&0\\0&\boldsymbol{\Delta}_j&0\\0&0&0\end{pmatrix},\tag{4}$$

上面的分块方式与 (3) 式的分块方式相同, $\boldsymbol{\Delta}_j$ 是 r_i 阶对角矩阵. 由 (2) 式知这是可以做到的. 则

$$\boldsymbol{P}_j\sum_{\substack{i=1\\i\neq j}}^{m}\boldsymbol{C}_i\boldsymbol{P}_j'=\boldsymbol{P}_j(\boldsymbol{I}-\boldsymbol{C}_j)\boldsymbol{P}_j'=\begin{pmatrix}\boldsymbol{I}&0&0\\0&\boldsymbol{I}-\boldsymbol{\Delta}_j&0\\0&0&\boldsymbol{I}\end{pmatrix}.\tag{5}$$

因为 (5) 式的秩不超过 (5) 式左上角和右下角$\boldsymbol{I}$的阶之和, 所以 $\boldsymbol{I}-\boldsymbol{\Delta}_j$ 的秩为 0, 即 $\boldsymbol{I}=\boldsymbol{\Delta}_j$. (因此 $\boldsymbol{C}_j$ 有 $\boldsymbol{r}_j$ 个非零特征根为 1, 则 $\boldsymbol{C}_j$ 是半正定矩阵.) 由 (4) 式可得

$$\boldsymbol{C}_j=\boldsymbol{P}_j'\begin{pmatrix}0&0&0\\0&\boldsymbol{I}&0\\0&0&0\end{pmatrix}\boldsymbol{P}_j=\boldsymbol{B}_j\boldsymbol{B}_j',\tag{6}$$

其中 $\boldsymbol{B}_j$ 由对应 (6) 式中 $\boldsymbol{I}$ 的 $\boldsymbol{P}_j'$ 的 r_j 列组成. 由 (1) 式可得

$$\boldsymbol{I}=\sum_{j=1}^{m}\boldsymbol{B}_j\boldsymbol{B}_j'=(\boldsymbol{B}_1,\boldsymbol{B}_2,\cdots,\boldsymbol{B}_m)\begin{pmatrix}\boldsymbol{B}_1'\\\boldsymbol{B}_2'\\\vdots\\\boldsymbol{B}_m'\end{pmatrix}=\boldsymbol{P}'\boldsymbol{P},\tag{7}$$

其中 $\boldsymbol{P}=(\boldsymbol{B}_1,\cdots,\boldsymbol{B}_m)'$. ■

下面给出 Cochran 定理的多元形式.

定理 7.4.1 设 $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_N$ 相互独立且服从 $N(0,\boldsymbol{\Sigma})$ 分布. 设矩阵 $(c_{\alpha\beta}^i)=\boldsymbol{C}_i$ 的秩为 r_i,

$$\boldsymbol{Q}_i=\sum_{\alpha,\beta=1}^{N}c_{\alpha\beta}^i\boldsymbol{Y}_\alpha\boldsymbol{Y}_\beta',\quad i=1,\cdots,m.\tag{8}$$

并且

$$\sum_{i=1}^{m}\boldsymbol{Q}_i=\sum_{\alpha=1}^{N}\boldsymbol{Y}_\alpha\boldsymbol{Y}_\alpha'.\tag{9}$$

则 (2) 式的一个充分必要条件为: $\boldsymbol{Q}_1,\cdots,\boldsymbol{Q}_m$ 相互独立, 且 $\boldsymbol{Q}_i$ 服从 $W(\boldsymbol{\Sigma},r_i)$ 分布.

由 (3) 式知 $\boldsymbol{C}_i$ 是幂等矩阵. 见附录的 A.2 节.

这个定理是单变量方差分析的一个非常有用的推广 (见第 8 章). 作为这个定理的一个应用, 我们将证明容量为 N 的样本均值与其转置的乘积具有奇异 Wishart 分布, 并且和具有非奇异 Wishart 分布的样本协方差阵的倍数是分布独立的. 设 $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_N$ 相互独立且服从 $N(0, \boldsymbol{\Sigma})$ 分布. 我们将用到矩阵 $\boldsymbol{C}_1 = (c_{\alpha\beta}^{(1)}) = (1/N)$ 和 $\boldsymbol{C}_2 = (c_{\alpha\beta}^{(2)}) = [\delta_{\alpha\beta} - (1/N)]$. 则

$$\boldsymbol{Q}_1 = \sum_{\alpha,\beta=1}^{N} \frac{1}{N} \boldsymbol{Y}_\alpha \boldsymbol{Y}'_\beta = N\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}', \tag{10}$$

$$\begin{aligned} \boldsymbol{Q}_2 &= \sum_{\alpha,\beta=1}^{N} (\delta_{\alpha\beta} - \frac{1}{N}) \boldsymbol{Y}_\alpha \boldsymbol{Y}'_\beta \\ &= \sum_{\alpha=1}^{N} \boldsymbol{Y}_\alpha \boldsymbol{Y}'_\alpha - N\bar{\boldsymbol{Y}}\bar{\boldsymbol{Y}}' \\ &= \sum_{\alpha=1}^{N} (\boldsymbol{Y}_\alpha - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_\alpha - \bar{\boldsymbol{Y}})', \end{aligned} \tag{11}$$

(9) 式也是成立的. 矩阵 $\boldsymbol{C}_1$ 的秩为 1, 矩阵 $\boldsymbol{C}_2$ 的秩为 $N-1$ (因为两个矩阵和的秩是小于或等于这两个矩阵秩的和, 又第二个矩阵的秩是小于 N 的). 定理中的条件都是满足的, 因此 $\boldsymbol{Q}_1$ 的分布与 $\boldsymbol{Z}\boldsymbol{Z}'$ 的分布相同, 其中 $\boldsymbol{Z}$ 服从 $N(0, \boldsymbol{\Sigma})$ 分布, 独立于 $\boldsymbol{Q}_2 \sim W(\boldsymbol{\Sigma}, N-1)$.

Anderson and Styan (1982) 对证明作了研究, 并扩展了 Cochran 定理.

7.5　广 义 方 差

7.5.1　广义方差的定义

单变量分布的方差 σ^2, 对应于多变量的协方差阵 $\boldsymbol{\Sigma}$. 另一种对应就是多变量的标量 $|\boldsymbol{\Sigma}|$, $|\boldsymbol{\Sigma}|$ 称为多元分布的*广义方差*[Wilks(1932); Frisch(1929)]. 类似地, 我们可以把样本向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 的广义方差定义为

$$|\boldsymbol{S}| = \left| \frac{1}{N-1} \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' \right|. \tag{1}$$

在某种意义上, 这两种方差都度量了数据的分散程度. 我们之所以在这里考虑它们, 是因为在假设检验中求似然比时经常会遇到样本广义方差.

样本广义方差的几何解释来自于把 $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$ 的 p 行看作 N 维空间中的 p 个向量. 由 3.2 节我们知道

$$(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}, \cdots, \boldsymbol{x}_N - \bar{\boldsymbol{x}}) = \boldsymbol{X} - \bar{\boldsymbol{x}}\boldsymbol{\varepsilon}' \tag{2}$$

的行正交于等角线 (过原点和ε), 其中 $\varepsilon=(1,\cdots,1)'$; 见图 3.2. 则

$$\boldsymbol{A}=(\boldsymbol{X}-\bar{\boldsymbol{x}}\boldsymbol{\varepsilon}')(\boldsymbol{X}-\bar{\boldsymbol{x}}\boldsymbol{\varepsilon}')' \tag{3}$$

的元素是 $\boldsymbol{X}-\bar{\boldsymbol{x}}\boldsymbol{\varepsilon}'$ 的行的内积.

我们现在定义一个由 n 维空间中的 p 个向量 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ 确定的超平行体 ($n\geqslant p$). 若 $p=1$, 超平行体是线段 $\boldsymbol{v}_1$. 若 $p=2$, 超平行体就是以 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 为主边的平行四边形. 也就是说, 它的边是 $\boldsymbol{v}_1,\boldsymbol{v}_2$, 平移 $\boldsymbol{v}_1$ 使得它的初始端点在 $\boldsymbol{v}_2$ 上, 平移 $\boldsymbol{v}_2$ 使得它的初始端点在 $\boldsymbol{v}_1$ 上. 见图 7.2. 若 $p=3$, 超平行体就是以 $\boldsymbol{v}_1,\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$ 为主边的平行六面体. 总之, 超平行体就是以 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ 为主边而定义的图形. 它被 p 对 $p-1$ 维的超平面截断, 每对中的两个超平面, 其中一个由 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ 中的 $p-1$ 个张成, 另外一个超平面穿过余下那个向量的端点.

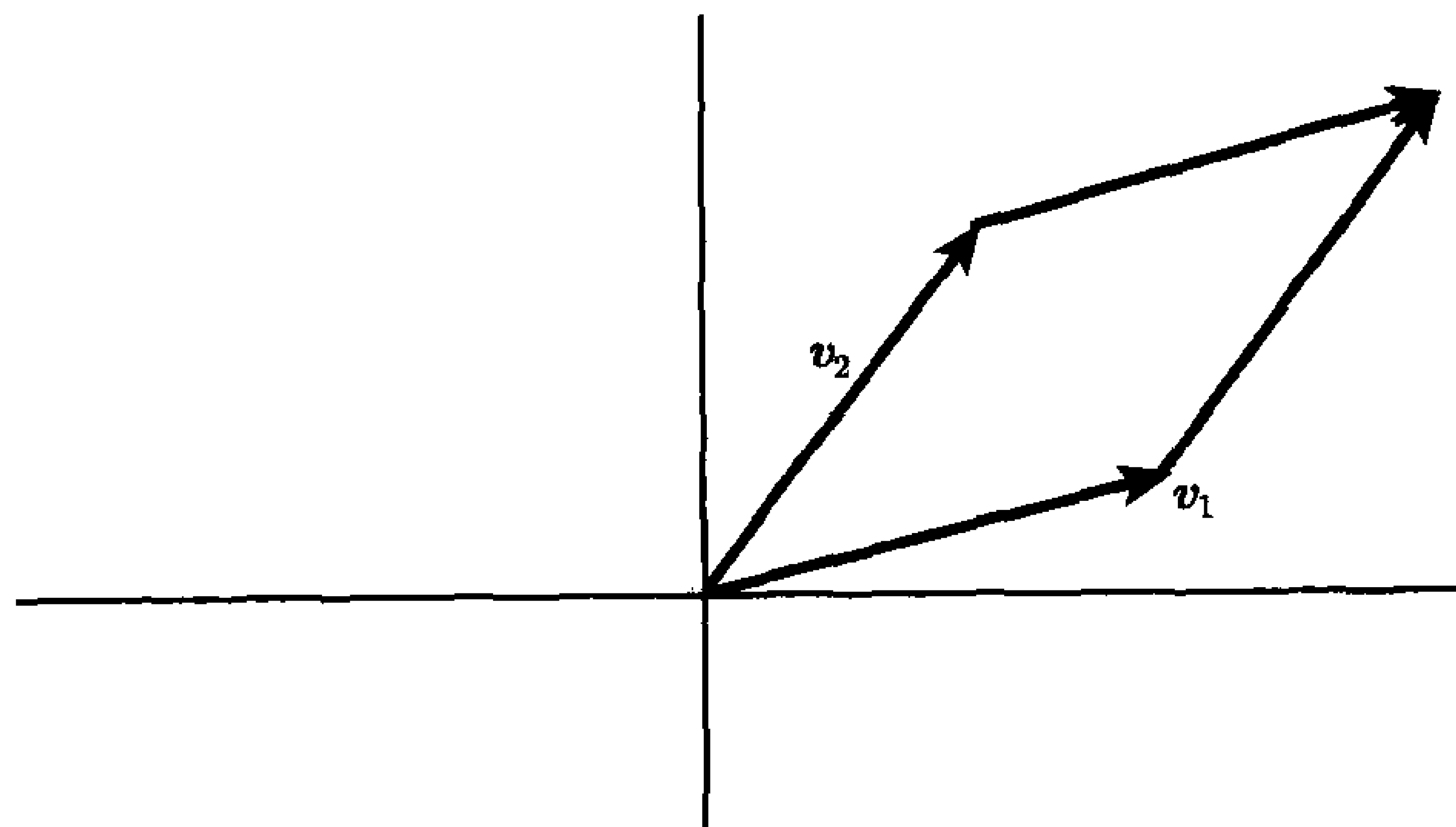


图 7.2 平行四边行

定理 7.5.1 *若 $\boldsymbol{V}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p)$, 则以 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p$ 为主边而构成的 p 维超平行体的体积的平方为 $|\boldsymbol{V}'\boldsymbol{V}|$.*

证明 若 $p=1$, 则 $|\boldsymbol{V}'\boldsymbol{V}|=\boldsymbol{v}_1'\boldsymbol{v}_1=\|\boldsymbol{v}_1\|^2$, 它就是 $\boldsymbol{v}_1$ 的一维体积的平方. 若两个 k 维的超平行体, 是由等体积等高的 $k-1$ 维的超平行体组成, 则这两个 k 维的超平行体的体积是相等的 [因为 k 维体积是通过 $k-1$ 维体积积分而得的]. 特别地, 若两个 k 维超平行体具有相同的 $k-1$ 维的超平行体和高度, 其中一个 k 维超平行体的第 k 个方向正交于前 $k-1$ 个方向, 则它们的体积相等. 因此, 以 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k$ 为主边的超平行体 (记为 P_k) 的体积等于以 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1}$ 为主边的超平行体 (记为 P_{k-1}) 的体积乘以 P_k 在 P_{k-1} 中的高度, 即

$$\mathrm{Vol}(P_k)=\mathrm{Vol}(P_{k-1})\times\mathrm{Alt}(P_k|P_{k-1}). \tag{4}$$

由前面的说明可得

$$\mathrm{Vol}(P_p)=\mathrm{Vol}(P_1)\times\mathrm{Alt}(P_2|P_1)\times\cdots\times\mathrm{Alt}(P_p|P_{p-1}). \tag{5}$$

由 7.2 节中的构造可知, P_k 在 P_{k-1} 中的高度为 $t_{kk}=\|\boldsymbol{w}_k\|$, 即 $\boldsymbol{v}_k$ 到由 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1}$

(或者 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{k-1}$) 张成的 $k-1$ 维空间的距离为 t_{kk}. 因此 $\mathrm{Vol}(P_p) = \prod_{k=1}^{p} t_{kk}$. 由 $|\boldsymbol{V}'\boldsymbol{V}| = |\boldsymbol{T}'\boldsymbol{T}| = \prod_{i=1}^{p} t_{ii}^2$ 知, 定理得证. ■

我们现在把该定理应用于以 (2) 的行作为主边的超平行体. 定理 7.5.1 中的维数可以是任意的 (但是至少为 p).

推论 7.5.1 以 (2) 的行作为主边的 p 维超平行体的体积的平方为 $|\boldsymbol{A}|$, 其中 $\boldsymbol{A}$ 由 (3) 式给出.

后面我们将会看到, 许多多元统计的知识都有基于这些体积的解释. 这些体积类似于特殊情况 $p=1$ 时的距离.

现在我们基于 p 维空间中的 N 个点来考虑 $|\boldsymbol{A}|$ 的几何解释. 设 (2) 的列 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ 表示 p 维空间的 N 个点. 当 $p=1$ 时, $|\boldsymbol{A}| = \sum_\alpha y_{1\alpha}^2$ 就是点到原点的距离的平方和. 总之, $|\boldsymbol{A}|$ 就是所有以集合 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ 中的 p 个向量为主边的超平行体的体积的平方和.

利用行列式展开的法则可得 [见附录中 A.1 节的 (24) 式]:

$$|\boldsymbol{A}| = \begin{vmatrix} \sum_\alpha y_{1\alpha}^2 & \cdots & \sum_\alpha y_{1\alpha}y_{p-1,\alpha} & \sum_\beta y_{1\beta}y_{p\beta} \\ \vdots & & \vdots & \vdots \\ \sum_\alpha y_{p-1,\alpha}y_{1\alpha} & \cdots & \sum_\alpha y_{p-1,\alpha}^2 & \sum_\beta y_{p-1,\beta}y_{p\beta} \\ \sum_\alpha y_{p\alpha}y_{1\alpha} & \cdots & \sum_\alpha y_{p\alpha}y_{p-1,\alpha} & \sum_\beta y_{p\beta}^2 \end{vmatrix} \tag{6}$$

$$= \sum_\beta \begin{vmatrix} \sum_\alpha y_{1\alpha}^2 & \cdots & \sum_\alpha y_{1\alpha}y_{p-1,\alpha} & y_{1\beta}y_{p\beta} \\ \vdots & & \vdots & \vdots \\ \sum_\alpha y_{p-1,\alpha}y_{1\alpha} & \cdots & \sum_\alpha y_{p-1,\alpha}^2 & y_{p-1,\beta}y_{p\beta} \\ \sum_\alpha y_{p\alpha}y_{1\alpha} & \cdots & \sum_\alpha y_{p\alpha}y_{p-1,\alpha} & y_{p\beta}^2 \end{vmatrix},$$

在 (6) 式中, 矩阵$\boldsymbol{A}$被分块为 $p-1$ 列和 1 列. 将行列式展开的法则连续应用于列, 可得

$$|\boldsymbol{A}| = \sum_{\alpha_1, \cdots, \alpha_p = 1}^{N} |y_{i\alpha_j} y_{j\alpha_j}|. \tag{7}$$

由定理 7.5.1 知, 以 $y_{\gamma_1}, \cdots, y_{\gamma_p}$ $(\gamma_1 < \cdots < \gamma_p)$ 为主边的超平行体的体积的平方为

$$\boldsymbol{V}_{\gamma_1, \cdots, \gamma_p}^2 = \left| \sum_\beta y_{i\beta} y_{j\beta} \right|, \tag{8}$$

其中对 β 的求和取遍 $(\gamma_1, \cdots, \gamma_p)$. 类似于 $|\boldsymbol{A}|$ 的方式, 我们也把这个行列式展开, 可得

$$V^2_{\gamma_1,\cdots,\gamma_p}=\sum|y_{i\beta_j}y_{j\beta_j}|, \tag{9}$$

其中对每个 β_j 的求和范围为 $(\gamma_1,\cdots,\gamma_p)$. 对于不同的 $\gamma_1<\cdots<\gamma_p$, 对 (9) 式求和, 得到 (7) 式. (若至少有两个 β_j 相等, 则 $|y_{i\beta_j}y_{j\beta_j}|=0$.) 因此 $|\boldsymbol{A}|$ 是由所有不同的以 y_α 中的 p 个向量作为主边的超平行体的体积的平方和. 用 $\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}}$ 替换 y_α, 得到下面的定理.

定理 7.5.2 设 $|\boldsymbol{S}|$ 由 (1) 式定义, 其中 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是 N 个向量组成的样本. 那么对于以 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 中的任意 p 个作为端点, 以 $\bar{\boldsymbol{x}}$ 作为另一个端点的 p 个向量, $|\boldsymbol{S}|$ 与以这 p 个向量为主边的不同超平行体的体积的平方和成比例, 比例因子为 $1/(N-1)^p$.

$|\boldsymbol{S}|$ 的总体模拟是 $|\boldsymbol{\Sigma}|$, 也可以给出 $|\boldsymbol{\Sigma}|$ 的几何解释. 由 3.3 节可知, 若 $\boldsymbol{X}$ 服从 $N(0,\boldsymbol{\Sigma})$ 分布, 则

$$\Pr\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\leqslant\chi^2_p(\alpha)\}=1-\alpha, \tag{10}$$

即 $\boldsymbol{X}$ 落入椭球

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}=\chi^2_p(\alpha) \tag{11}$$

内的概率为 $1-\alpha$. 该椭球的体积为 $C(p)|\boldsymbol{\Sigma}|^{\frac{1}{2}}[\chi^2_p(\alpha)]^{\frac{1}{2}p}/p$, 其中 $C(p)$ 在习题 7.3 中定义.

7.5.2 样本广义方差的分布

$|\boldsymbol{S}|$ 的分布与 $|\boldsymbol{A}|/(N-1)^p$ 的分布相同, 其中 $\boldsymbol{A}=\sum_{\alpha=1}^n\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha$, $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 相互独立且服从 $N(0,\boldsymbol{\Sigma})$ 分布, $n=N-1$. 设 $\boldsymbol{Z}_\alpha=\boldsymbol{C}\boldsymbol{Y}_\alpha$, $\alpha=1,\cdots,n$, 其中 $\boldsymbol{C}\boldsymbol{C}'=\boldsymbol{\Sigma}$. 则 $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_n$ 相互独立且都服从 $N(0,\boldsymbol{I})$ 分布. 设

$$\boldsymbol{B}=\sum_{\alpha=1}^n\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\alpha=\sum_{\alpha=1}^n\boldsymbol{C}^{-1}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha(\boldsymbol{C}^{-1})'=\boldsymbol{C}^{-1}\boldsymbol{A}(\boldsymbol{C}^{-1})', \tag{12}$$

则 $|\boldsymbol{A}|=|\boldsymbol{C}|\cdot|\boldsymbol{B}|\cdot|\boldsymbol{C}'|=|\boldsymbol{B}|\cdot|\boldsymbol{\Sigma}|$. 推广 7.2 节的结果可知, $|\boldsymbol{B}|$ 的分布与 $\prod_{i=1}^p t_{ii}^2$ 的分布相同, $t_{11}^2,\cdots,t_{pp}^2$ 相互独立且都服从 χ^2 分布.

定理 7.5.3 来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的样本 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 的广义方差 $|\boldsymbol{S}|$ 的分布, 和 $|\boldsymbol{\Sigma}|/(N-1)^p$ 与 p 个独立因子乘积的分布相同, 其中第 i 个因子服从自由度为 $N-i$ 的 χ^2 分布.

若 $p=1$, 则 $|\boldsymbol{S}|$ 的分布与 $|\boldsymbol{\Sigma}|\cdot\chi^2_{N-1}/(N-1)$ 的分布相同. 若 $p=2$, $|\boldsymbol{S}|$ 的分布与 $|\boldsymbol{\Sigma}|\cdot\chi^2_{N-1}\cdot\chi^2_{N-2}/(N-1)^2$ 的分布相同. 由习题 7.15 或者习题 7.37 可知, 当 $p=2$ 时, $|\boldsymbol{S}|$ 的分布与 $|\boldsymbol{\Sigma}|\cdot(\chi^2_{2N-4})^2/(2N-2)^2$ 的分布相同. 因为

$$|\boldsymbol{A}|=|\boldsymbol{\Sigma}|\times\chi^2_{N-1}\times\chi^2_{N-2}\times\cdots\times\chi^2_{N-p}, \tag{13}$$

所以当 $p=2r$ 时 $|\boldsymbol{A}|$ 的分布与

$$|\boldsymbol{\Sigma}|(\chi^2_{2N-4}\times\chi^2_{2N-8}\times\cdots\times\chi^2_{2N-4r})^2/2^{2r} \tag{14}$$

的分布相同. 因为自由度为 m 的 χ^2 变量的 h 阶矩为 $2^h\Gamma(\frac{1}{2}m+h)/\Gamma(\frac{1}{2}m)$, 又独立变量积的矩等于变量矩的积, 则 $|\boldsymbol{A}|$ 的 h 阶矩为

$$|\boldsymbol{\Sigma}|^h\prod_{i=1}^{p}\left\{2^h\frac{\Gamma[\frac{1}{2}(N-i)+h]}{\Gamma[\frac{1}{2}(N-i)]}\right\}=2^{hp}|\boldsymbol{\Sigma}|^h\frac{\prod_{i=1}^{p}\Gamma[\frac{1}{2}(N-i)+h]}{\prod_{i=1}^{p}\Gamma[\frac{1}{2}(N-i)]} \tag{15}$$
$$=2^{hp}|\boldsymbol{\Sigma}|^h\frac{\Gamma_p[\frac{1}{2}(N-1)+h]}{\Gamma_p[\frac{1}{2}(N-1)]}.$$

因此

$$\mathrm{E}|\boldsymbol{A}|=|\boldsymbol{\Sigma}|\prod_{i=1}^{p}(N-i). \tag{16}$$

$$\mathrm{Var}(|\boldsymbol{A}|)=|\boldsymbol{\Sigma}|^2\prod_{i=1}^{p}(N-i)\left[\prod_{j=1}^{p}\Gamma(N-j+2)-\prod_{j=1}^{p}\Gamma(N-j)\right], \tag{17}$$

其中 $\mathrm{Var}(|\boldsymbol{A}|)$ 为 $|\boldsymbol{A}|$ 的方差.

7.5.3 样本广义方差的渐近分布

设 $|\boldsymbol{B}|/n^p=V_1(n)\times V_2(n)\times\cdots\times V_p(n)$, 其中 $V_i(n)$ 分布独立且 $nV_i(n)=\chi^2_{n-p+i}$. 因为 χ^2_{n-p+i} 的分布与 $\sum_{\alpha=1}^{n-p+i}W_\alpha^2$ 的分布是相同的, 其中 W_α 相互独立且服从 $N(0,1)$ 分布, 所以由中心极限定理知 (应用于 W_α^2)

$$\frac{nV_i(n)-(n-p+i)}{\sqrt{2(n-p+i)}}=\sqrt{n}\frac{V_i(n)-1+\frac{p-i}{n}}{\sqrt{2}\sqrt{1-\frac{p-i}{n}}} \tag{18}$$

的渐近分布为 $N(0,1)$. 因此 $\sqrt{n}[V_i(n)-1]$ 的渐近分布为 $N(0,2)$. 应用定理 4.2.3 可得

$$\boldsymbol{U}(n)=\begin{pmatrix}V_1(n)\\ \vdots\\ V_p(n)\end{pmatrix},\quad \boldsymbol{b}=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}, \tag{19}$$

$|\boldsymbol{B}|/n^p=w=f(u_1,\cdots,u_p)=u_1u_2\cdots u_p$, $\boldsymbol{T}=2\boldsymbol{I}, \partial f\Big/\partial u_i|_{\boldsymbol{u}=\boldsymbol{b}}=1, \phi_{\boldsymbol{b}}'\boldsymbol{T}\phi_{\boldsymbol{b}}=2p$.
因此

$$\sqrt{n}(|\boldsymbol{B}|/n^p-1) \tag{20}$$

的渐近分布为 $N(0,2p)$.

定理 7.5.4 设$\boldsymbol{S}$是自由度为 n 的 $p\times p$ 的样本协方差阵. 则 $\sqrt{n}(|\boldsymbol{S}|/|\boldsymbol{\Sigma}|-1)$ 的渐近分布是均值为 0 方差为 $2p$ 的正态分布.

7.6 总体协方差阵为对角矩阵时相关系数集的分布

4.2.1 节给出了总体的相关系数为 0 时单个样本相关系数的分布. 下面我们来求当 $\rho_{ij}=0\,(i<j)$ 时, 集合 $r_{ij}\,(i<j,i=1,\cdots,p,j=1,\cdots,p)$ 的密度.

首先来讨论当$\boldsymbol{\Sigma}$为对角矩阵时$\boldsymbol{A}$的分布, 即 $W[(\sigma_{ii}\delta_{ij}),n]$. 由

$$|\boldsymbol{\Sigma}| = \begin{vmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{vmatrix} = \prod_{i=1}^{p} \sigma_{ii} \tag{1}$$

知, $\boldsymbol{A}$ 的密度为

$$\frac{|a_{ij}|^{\frac{1}{2}(n-p-1)} \exp(-\frac{1}{2}\sum_{i=1}^{p} a_{ii}/\sigma_{ii})}{2^{\frac{1}{2}np}\prod_{i=1}^{p}\sigma_{ii}^{\frac{1}{2}n}\Gamma_p(\frac{1}{2}n)}. \tag{2}$$

做变换

$$a_{ij} = \sqrt{a_{ii}}\sqrt{a_{jj}}r_{ij}, \quad i \neq j, \tag{3}$$

$$a_{ii} = a_{ii}. \tag{4}$$

对于固定的 a_{ii}, 变换的雅可比行列式为 (3) 的雅可比行列式和 (4) 的雅可比行列式之积. (3) 的雅可比行列式为对角元为 $\sqrt{a_{ii}}\sqrt{a_{jj}}$ 的 $p(p-1)/2$ 阶对角矩阵的行列式. 因为每一个固定的下标 k 在集合 $r_{ij}(i<j)$ 中出现 $p-1$ 次, 所以雅可比行列式为

$$\boldsymbol{J} = \prod_{i=1}^{p} a_{ii}^{\frac{1}{2}(p-1)}. \tag{5}$$

如果把 (3) 和 (4) 代入 $w[\boldsymbol{A}|(\sigma_{ii}\delta_{ij}),n]$ 并乘以 (5), 再由

$$|\sqrt{a_{ii}}\sqrt{a_{jj}}r_{ij}| = \left(\prod_{i=1}^{p} a_{ii}\right)|r_{ij}|, \tag{6}$$

可得 $\{a_{ii}\}$ 和 $\{r_{ii}\}$ 的联合密度为

$$\frac{|\sqrt{a_{ii}}\sqrt{a_{jj}}r_{ij}|^{\frac{1}{2}(n-p-1)} \exp(-\frac{1}{2}\sum_{i=1}^{p} a_{ii}/\sigma_{ii})}{2^{\frac{1}{2}np}\prod_{i=1}^{p}\sigma_{ii}^{\frac{1}{2}n}\Gamma_p(\frac{1}{2}n)} \prod_{i=1}^{p} a_{ii}^{\frac{1}{2}(p-1)} \tag{7}$$

$$= \frac{|r_{ij}|^{\frac{1}{2}(n-p-1)}}{\Gamma_p(\frac{1}{2}n)} \prod_{i=1}^{p} \left\{ \frac{a_{ii}^{\frac{1}{2}n-1}\exp(-\frac{1}{2}a_{ii}/\sigma_{ii})}{2^{\frac{1}{2}n}\sigma_{ii}^{\frac{1}{2}n}} \right\},$$

其中 $r_{ii}=1$. 在 (6) 式右端乘积的第 i 项中, 令 $a_{ii}/(2\sigma_{ii}) = u_i$, 则由伽玛函数的定义 (也可由 a_{ii}/σ_{ii} 服从自由度为 n 的 χ^2 分布) 可得该项积分为

$$\int_0^{\infty} \frac{a_{ii}^{\frac{1}{2}n-1}\exp(-\frac{1}{2}a_{ii}/\sigma_{ii})}{2^{\frac{1}{2}n}\sigma_{ii}^{\frac{1}{2}n}} \mathrm{d}a_{ii} = \int_0^{\infty} u_i^{\frac{1}{2}n-1}\mathrm{e}^{-u_i}\mathrm{d}u_i = \Gamma\left(\frac{1}{2}n\right). \tag{8}$$

从而 r_{ij} 的密度为

$$\frac{\Gamma^p(\frac{1}{2}n)|r_{ij}|^{\frac{1}{2}(n-p-1)}}{\Gamma_p(\frac{1}{2}n)}. \tag{9}$$

定理 7.6.1 若 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 相互独立而且服从 $N[\boldsymbol{\mu},(\sigma_{ii}\delta_{ij})]$ 分布, 则 (9) 式为样本相关系数的密度, 其中 $n=N-1$.

7.7 逆 Wishart 分布, 协方差阵的贝叶斯估计

7.7.1 逆 Wishart 分布

由 3.4.2 节可知, 一般情况下, 贝叶斯估计是容许的. 而且参数的先验分布选择得合适时, 贝叶斯估计是易于计算的. 如果存在一个充分统计量, 则存在参数的一个先验分布类, 使得它的后验分布也属于这个分布类; 这样的类称为**共轭分布类**. 由 3.4.2 节可知, 当协方差阵给定时, 先验的正态类共轭于正态分布类. 在这一节我们将讨论协方差阵的贝叶斯估计以及均值向量与协方差阵的估计.

定理 7.7.1 若 $\boldsymbol{A}$ 的分布为 $W(\boldsymbol{\Sigma},m)$, 则当 $\boldsymbol{B}=\boldsymbol{A}^{-1}$ 正定时, $\boldsymbol{B}$ 的密度为

$$\frac{|\boldsymbol{\Psi}|^{\frac{1}{2}m}|\boldsymbol{B}|^{-\frac{1}{2}(m+p+1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Psi}\boldsymbol{B}^{-1}}}{2^{\frac{1}{2}mp}\Gamma_p(\frac{1}{2}m)}, \tag{1}$$

否则为 0, 其中 $\boldsymbol{\Psi}=\boldsymbol{\Sigma}^{-1}$.

证明 由附录的定理 A.4.6 可知, 变换 $\boldsymbol{A}=\boldsymbol{B}^{-1}$ 的雅可比行列式为 $|\boldsymbol{B}|^{-(p+1)}$. 用 $\boldsymbol{B}^{-1}$ 替换 7.2 节 (14) 式中的$\boldsymbol{A}$, 然后再乘以 $|\boldsymbol{B}|^{-(p+1)}$, 结论可证.

我们把 (1) 称为自由度[①]为 m 的逆 Wishart 分布, 并记为 $W^{-1}(\boldsymbol{\Psi},m)$, 其密度记为 $w^{-1}(\boldsymbol{B}|\boldsymbol{\Psi},m)$. 称 $\boldsymbol{\Psi}$ 为精度 (precision) 矩阵或集中 (concentration) 矩阵.

7.7.2 协方差阵的贝叶斯估计

来自 $N(\boldsymbol{\mu},\boldsymbol{\sigma})$ 的样本量为 N 的协方差阵的分布为 $(1/n)\boldsymbol{A}$, 其中 $\boldsymbol{A}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布, $n=N-1$. 我们下面将说明当 $\boldsymbol{\Sigma}$ 服从逆 Wishart 分布时, 给定 $\boldsymbol{A}$ 时 $\boldsymbol{\Sigma}$ 的条件分布也是逆 Wishart 分布, 即 $\boldsymbol{\Sigma}$ 的逆 Wishart 分布类共轭于 Wishart 分布类.

定理 7.7.2 若 $\boldsymbol{A}$ 的分布为 $W(\boldsymbol{\Sigma},n)$, 且 $\boldsymbol{\Sigma}$ 的先验分布为 $W^{-1}(\boldsymbol{\Psi},m)$, 则 $\boldsymbol{\Sigma}$ 的条件分布为 $W^{-1}(\boldsymbol{A}+\boldsymbol{\Psi},n+m)$.

证明 对于正定矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{\Sigma}$, 其联合分布为

$$\frac{|\boldsymbol{\Psi}|^{\frac{1}{2}m}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(n+m+p+1)}|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{A}+\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}}}{2^{\frac{1}{2}(n+m)p}\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}m)}. \tag{2}$$

① 自由度的定义与 Giri (1977) 的 104 页和 Muirhead (1982) 的 113 页的定义不同.

对 (2) 式在所有正定矩阵 $\boldsymbol{\Sigma}$ 的集合上进行积分, 可得 $\boldsymbol{A}$ 的边际分布. 因为对于一个$\boldsymbol{\Psi}$, (1) 式关于 $\boldsymbol{B}$ 的积分值为 1, 则对于正定的 $\boldsymbol{A}$, (2) 式关于 $\boldsymbol{\Sigma}$ 积分为

$$\frac{\Gamma_p[\frac{1}{2}(n+m)]|\boldsymbol{\Psi}|^{\frac{1}{2}m}|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}|\boldsymbol{A}+\boldsymbol{\Psi}|^{-\frac{1}{2}(n+m)}}{\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}m)}. \tag{3}$$

给定 $\boldsymbol{A}$ 时 $\boldsymbol{\Sigma}$ 的条件密度为 (2) 和 (3) 的比,

$$\frac{|\boldsymbol{A}+\boldsymbol{\Psi}|^{\frac{1}{2}(n+m)}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(n+m+p+1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{A}+\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}}}{2^{\frac{1}{2}(n+m)p}\Gamma_p[\frac{1}{2}(n+m)]}, \tag{4}$$

即 $w^{-1}(\boldsymbol{\Sigma}|\boldsymbol{A}+\boldsymbol{\Psi},n+m)$. ■

推论 7.7.1 如果 $n\boldsymbol{S}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布, 且 $\boldsymbol{\Sigma}$ 的先验分布为 $W^{-1}(\boldsymbol{\Psi},m)$, 则给定 $\boldsymbol{S}$ 时 $\boldsymbol{\Sigma}$ 的条件分布为 $W^{-1}(n\boldsymbol{S}+\boldsymbol{\Psi},n+m)$.

推论 7.7.2 如果 $n\boldsymbol{S}$ 服从 $W(\boldsymbol{\Sigma},n)$ 分布, 且 $\boldsymbol{\Sigma}$ 的先验分布为 $W^{-1}(\boldsymbol{\Psi},m)$, 损失函数为 $\mathrm{tr}(\boldsymbol{D}-\boldsymbol{\Sigma})\boldsymbol{G}(\boldsymbol{D}-\boldsymbol{\Sigma})\boldsymbol{H}$, 其中 $\boldsymbol{G}$ 和 $\boldsymbol{H}$ 为正定矩阵, 则 $\boldsymbol{\Sigma}$ 的贝叶斯估计为

$$\frac{1}{n+m-p-1}(n\boldsymbol{S}+\boldsymbol{\Psi}). \tag{5}$$

证明 由 3.4.2 节知, $\boldsymbol{\Sigma}$ 的贝叶斯估计为 $\mathrm{E}(\boldsymbol{\Sigma}|\boldsymbol{S})$. 由定理 7.7.2 可知, $\boldsymbol{\Sigma}^{-1}$ 的后验分布为 $W[(n\boldsymbol{S}+\boldsymbol{\Psi})^{-1},n+m]$. 由下面的引理可得该定理. ■

引理 7.7.1 若$\boldsymbol{A}$服从 $W(\boldsymbol{\Sigma},n)$ 分布, 则

$$\mathrm{E}(\boldsymbol{A}^{-1})=\frac{1}{n-p-1}\boldsymbol{\Sigma}^{-1}. \tag{6}$$

证明 若 $\boldsymbol{C}$ 是一个非奇异矩阵且满足 $\boldsymbol{\Sigma}=\boldsymbol{C}\boldsymbol{C}'$, 则 $\boldsymbol{A}$ 与 $\boldsymbol{\Sigma}=\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}'$ 的分布相同, 其中 $\boldsymbol{B}$ 的分布为 $W(\boldsymbol{I},n)$, 且 $\mathrm{E}(\boldsymbol{A}^{-1})=(\boldsymbol{C}')^{-1}(\mathrm{E}(\boldsymbol{B}^{-1}))\boldsymbol{C}^{-1}$. 由对称性知, $\mathrm{E}(\boldsymbol{B}^{-1})$ 的对角元都相同, 非对角元也相同, 即 $\mathrm{E}(\boldsymbol{B}^{-1})=k_1\boldsymbol{I}+k_2\varepsilon\varepsilon'$. 给定任意一个正交矩阵 $\boldsymbol{Q}$, $\boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}'$ 服从 $W(\boldsymbol{I},n)$ 分布, 则 $\mathrm{E}(\boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}')^{-1}=\boldsymbol{Q}\mathrm{E}(\boldsymbol{B}^{-1})\boldsymbol{Q}'=\mathrm{E}(\boldsymbol{B}^{-1})$. 因此 $k_2=0$. $\boldsymbol{B}^{-1}$ 的一个对角元的分布为 $(\chi^2_{n-p-1})^{-1}$ (见定理 5.2.2 的证明). 由 $\mathrm{E}(\chi^2_{n-p+1})^{-1}=(n-p-1)^{-1}$, $\mathrm{E}(\boldsymbol{B}^{-1})=(n-p-1)^{-1}\boldsymbol{I}$, 可得 (6) 式. ■

注意 $(n-p-1)\boldsymbol{A}^{-1}=[(n-p-1)/(n-1)]\boldsymbol{S}^{-1}$ 是精度 $\boldsymbol{\Sigma}^{-1}$ 的一个无偏估计.

若 $\boldsymbol{\mu}$ 已知, $\boldsymbol{\Sigma}$ 的无偏估计为 $(1/N)\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha-\boldsymbol{\mu})(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'$. 以上可以用 N 代替 n. 另外, 若 n (或 N) 比较大时, (5) 式近似于 $\boldsymbol{S}$.

定理 7.7.3 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测. 并设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 具有先验密度 $n(\boldsymbol{\mu}|\boldsymbol{\nu},(1/K)\boldsymbol{\Sigma})\times w^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Psi},m)$. 则给定 $\bar{\boldsymbol{x}}=(1/N)\sum_{\alpha=1}^N\boldsymbol{x}_\alpha$ 和 $\boldsymbol{S}=(1/n)\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$ 时, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的后验密度为

$$n\left(\boldsymbol{\mu}\middle|\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu}),\frac{1}{N+K}\boldsymbol{\Sigma}\right) \tag{7}$$

$$\cdot w^{-1}\left(\boldsymbol{\Sigma}\middle|n\boldsymbol{S}+\boldsymbol{\Psi}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})',N+m\right).$$

证明 因为 $\bar{\boldsymbol{x}}$ 和 $n\boldsymbol{S}=\boldsymbol{A}$ 是充分统计量, 考虑 $\bar{\boldsymbol{x}},\boldsymbol{A},\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的联合密度,

$$\frac{K^{\frac{1}{2}p}N^{\frac{1}{2}p}|\boldsymbol{\Psi}|^{\frac{1}{2}m}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+m+p+2)}|\boldsymbol{A}|^{\frac{1}{2}(N-p-2)}}{2^{\frac{1}{2}(N+m+1)p}\pi^{p}\Gamma_p[\frac{1}{2}(N-1)]\Gamma_p(\frac{1}{2}m)} \cdot\exp\Big\{-\frac{1}{2}[N(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu})+\mathrm{tr}\boldsymbol{A}\boldsymbol{\Sigma}^{-1} +K(\boldsymbol{\mu}-\boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\nu})+\mathrm{tr}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}]\Big\} \tag{8}$$

对 (8) 式关于 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 积分可得 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}$ 的边际密度. (8) 式的指数是 $-\frac{1}{2}$ 乘以

$$\begin{aligned}&(N+K)\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-2(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\\&\qquad+N\bar{\boldsymbol{x}}'\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}+K\boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}+\mathrm{tr}(\boldsymbol{A}+\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}\\&=(N+K)\left[\boldsymbol{\mu}-\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{\mu}-\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu})\right]\\&\quad+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})+\mathrm{tr}(\boldsymbol{A}+\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}.\end{aligned} \tag{9}$$

对 (8) 式关于 $\boldsymbol{\mu}$ 积分可得

$$\frac{K^{\frac{1}{2}p}N^{\frac{1}{2}p}|\boldsymbol{\Psi}|^{\frac{1}{2}m}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+m+p+1)}|\boldsymbol{A}|^{\frac{1}{2}(N-p-2)}}{(N+K)^{\frac{1}{2}p}2^{\frac{1}{2}(N+m)p}\pi^{\frac{1}{2}p}\Gamma_p[\frac{1}{2}(N-1)]\Gamma_p(\frac{1}{2}m)} \cdot\exp\left\{-\frac{1}{2}\left[\mathrm{tr}\boldsymbol{A}\boldsymbol{\Sigma}^{-1}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})+\mathrm{tr}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right]\right\}. \tag{10}$$

然后对 (10) 式关于 $\boldsymbol{\Sigma}$ 积分可得

$$\frac{K^{\frac{1}{2}p}N^{\frac{1}{2}}\Gamma_p[\frac{1}{2}(N+m)]}{\pi^{\frac{1}{2}p}\Gamma_p[\frac{1}{2}(N-1)]\Gamma_p(\frac{1}{2}m)(N+K)^{\frac{1}{2}p}} \cdot|\boldsymbol{A}|^{\frac{1}{2}(N-p-2)}|\boldsymbol{\Psi}|^{\frac{1}{2}m}\left|\boldsymbol{\Psi}+\boldsymbol{A}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\right|^{-\frac{1}{2}(N+m)}. \tag{11}$$

给定 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{A}$ 时, $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的条件密度为 (8) 式与 (11) 式的比, 即

$$\begin{aligned}&\frac{(N+K)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{-\frac{1}{2}(N+m+p+2)}|\boldsymbol{\Psi}+\boldsymbol{A}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'|^{\frac{1}{2}(N+m)}}{2^{\frac{1}{2}(N+m+1)p}\pi^{\frac{1}{2}p}\Gamma_p[\frac{1}{2}(N+m)]}\\&\cdot\exp\left\{-\frac{1}{2}\left[(N+K)\left[\boldsymbol{\mu}-\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu})\right]'\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\mu}-\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu})\right]\right.\right.\\&\left.\left.+\mathrm{tr}\left[\boldsymbol{\Psi}+\boldsymbol{A}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\right]\boldsymbol{\Sigma}^{-1}\right]\right\}.\end{aligned} \tag{12}$$

则 (12) 式可以化为 (7) 式. ■

推论 7.7.3 若 $x_1,\cdots,x_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测. 并设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 具有先验密度 $n(\boldsymbol{\mu}|\boldsymbol{\nu},(1/K)\boldsymbol{\Sigma})\times w^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Psi},m)$, 损失函数为 $(\boldsymbol{d}-\boldsymbol{\mu})'\boldsymbol{J}(\boldsymbol{d}-\boldsymbol{\mu})-\mathrm{tr}(\boldsymbol{D}-\boldsymbol{\Sigma})\boldsymbol{G}(\boldsymbol{D}-\boldsymbol{\Sigma})\boldsymbol{H}$, 则 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的贝叶斯估计分别为

$$\frac{1}{N+K}(N\bar{\boldsymbol{x}}+K\boldsymbol{\nu}), \tag{13}$$

$$\frac{1}{N+m-p-1}\left[n\boldsymbol{S}+\boldsymbol{\Psi}+\frac{NK}{N+K}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\right]. \tag{14}$$

$\boldsymbol{\mu}$ 的估计为样本均值 $\bar{\boldsymbol{x}}$ 和先验均值 $\boldsymbol{\nu}$ 的加权平均. 当 N 比较大时先验均值的权比较小. $\boldsymbol{\Sigma}$ 的估计是协方差阵 $\boldsymbol{S}$, $\boldsymbol{\Psi}$ 以及样本均值与先验均值差的加权平均. 当 N 比较大时估计比较接近样本协方差阵.

定理 7.7.4 设 $x_1,\cdots,x_N$ 是来自 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的观测. 并设 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 具有先验密度 $n[\boldsymbol{\mu}|\boldsymbol{\nu},(1/K)\boldsymbol{\Sigma}]\times w^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Psi},m)$. 则给定 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 时, $\boldsymbol{\mu}$ 的后验边际密度为

$$\frac{(N+K)^{\frac{1}{2}p}\Gamma[\frac{1}{2}(N+m+1)]|\boldsymbol{B}|^{-\frac{1}{2}}}{\pi^{\frac{1}{2}p}\Gamma[\frac{1}{2}(N+m+1-p)][1+(N+K)(\boldsymbol{\mu}-\boldsymbol{\mu}^*)'\boldsymbol{B}^{-1}(\boldsymbol{\mu}-\boldsymbol{\mu}^*)]^{\frac{1}{2}(N+m-1)}}, \tag{15}$$

其中 $\boldsymbol{\mu}^*$ 是 (13), $\boldsymbol{B}$ 是 $N+m-p-1$ 与 (14) 的积.

证明 (12) 式的指数是 $-\frac{1}{2}$ 乘以

$$\mathrm{tr}[\boldsymbol{B}+(N+K)(\boldsymbol{\mu}-\boldsymbol{\mu}^*)(\boldsymbol{\mu}-\boldsymbol{\mu}^*)']\boldsymbol{\Sigma}^{-1}. \tag{16}$$

由 $|\boldsymbol{B}+\boldsymbol{x}\boldsymbol{x}'|=|\boldsymbol{B}|(1+\boldsymbol{x}'\boldsymbol{B}^{-1}\boldsymbol{x})$(推论 A.3.1) 可知, (12) 式关于 $\boldsymbol{\Sigma}$ 积分可得

$$\frac{(N+K)^{\frac{1}{2}p}\Gamma_p[\frac{1}{2}(N+m+1)]|\boldsymbol{B}|^{\frac{1}{2}(N+m)}}{\pi^{\frac{1}{2}p}\Gamma_p[\frac{1}{2}(N+m)]|\boldsymbol{B}+(N+K)(\boldsymbol{\mu}-\boldsymbol{\mu}^*)(\boldsymbol{\mu}-\boldsymbol{\mu}^*)'|^{\frac{1}{2}(N+m-1)}}, \tag{17}$$

从而可得 (15) 式. ■

(15) 的密度是自由度为 $N+m+1-p$ 的多元 t 分布. 见 2.7.5 节的例子.

7.8 协方差阵的改进估计

当损失函数为二次的时候, 样本均值 $\bar{\boldsymbol{x}}$ 作为总体均值 $\boldsymbol{\mu}$ 的估计可以被改进, 类似地, 对一定的损失函数, 样本协方差阵 $\boldsymbol{S}$ 作为总体协方差阵 $\boldsymbol{\Sigma}$ 的估计也可以被改进. 位置参数 $\boldsymbol{\mu}$ 的估计的损失函数是关于变换 $(\boldsymbol{x}\to\boldsymbol{x}+\boldsymbol{a},\ \boldsymbol{\mu}\to\boldsymbol{\mu}+\boldsymbol{a})$ 不变的, 样本均值 $\boldsymbol{\mu}$ (当 $\boldsymbol{\Sigma}$ 已知时, 它是唯一的无偏的充分统计量的函数) 的风险不依赖于参数值. 协方差阵的一组自然的变换就是在其左边乘以一个非奇异矩阵, 在其右边乘以该非奇异矩阵的转置 $(\boldsymbol{x}\to\boldsymbol{C}\boldsymbol{x},\ \boldsymbol{S}\to\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}',\ \boldsymbol{\Sigma}\to\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}')$. 我们考虑两类关于这种变换不变的损失函数.

一类损失函数为二次的,

$$L_q(\boldsymbol{\Sigma}, \boldsymbol{G}) = \operatorname{tr}(\boldsymbol{G} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}(\boldsymbol{G} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1} \tag{1}$$
$$= \operatorname{tr}(\boldsymbol{G}\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})^2,$$

其中 $\boldsymbol{G}$ 为一正定矩阵. 另一类基于似然函数的形式 (见引理 3.2.2, 以及在习题 3.4、习题 3.8 和习题 3.12 中的另一种证明):

$$L_l(\boldsymbol{\Sigma}, \boldsymbol{G}) = \operatorname{tr}\boldsymbol{G}\boldsymbol{\Sigma}^{-1} - \ln|\boldsymbol{G}\boldsymbol{\Sigma}^{-1}| - p. \tag{2}$$

当 $\boldsymbol{G} \neq \boldsymbol{\Sigma}$ 时, 这两个损失函数都是正的, 否则为 0. 当$\boldsymbol{G}$接近于奇异矩阵或者 $\boldsymbol{G}$ 有一个或多个元素 (一个或多个特征根) 接近于 ∞ 时, 第二个损失函数接近于 ∞(见引理 3.2.2 的证明). 这两个损失函数对变换 $\boldsymbol{G}^* = \boldsymbol{C}\boldsymbol{G}\boldsymbol{C}'$, $\boldsymbol{\Sigma}^* = \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}'$ 是不变的. 从 $L_q(\boldsymbol{I}, \boldsymbol{D}) = \sum_{i=1}^{p}(d_{ii} - 1)^2$ 和 $L_l(\boldsymbol{I}, \boldsymbol{D}) = \sum_{i=1}^{p}(d_{ii} - \ln d_{ii} - 1)$ 中可以看出损失函数的一些性质, 其中 $\boldsymbol{D}$ 是对角矩阵. (由附录中的定理 A.2.2 知, 对于任意的正定矩阵 $\boldsymbol{\Sigma}$ 和对称矩阵 $\boldsymbol{G}$, 存在一个非奇异矩阵 $\boldsymbol{C}$ 满足 $\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}' = \boldsymbol{I}$ 和 $\boldsymbol{C}\boldsymbol{G}\boldsymbol{C}' = \boldsymbol{D}$.) 如果设 $\boldsymbol{g} = (g_{11}, \cdots, g_{pp}, g_{12}, \cdots, g_{p-1,p})'$, $\boldsymbol{s} = (s_{11}, \cdots, s_{pp}, s_{12}, \cdots, s_{p-1,p})'$, $\boldsymbol{\sigma} = (\sigma_{11}, \cdots, \sigma_{pp}, \sigma_{12}, \cdots, \sigma_{p-1,p})'$, $\boldsymbol{\Phi} = \mathrm{E}(\boldsymbol{s} - \boldsymbol{\sigma})(\boldsymbol{s} - \boldsymbol{\sigma})'$, 那么 $L_q(\boldsymbol{\Sigma}, \boldsymbol{G})$ 是 $(\boldsymbol{g} - \boldsymbol{\sigma})'\boldsymbol{\Phi}^{-1}(\boldsymbol{g} - \boldsymbol{\sigma})$ 的常数倍 (见问题 7.3.3).

极大似然估计 $\hat{\boldsymbol{\Sigma}}$ 和无偏估计 $\boldsymbol{S}$ 都具有形式 $a\boldsymbol{A}$, 其中 $\boldsymbol{A}$ 有分布 $W(\boldsymbol{\Sigma}, n), n = N - 1$.

定理 7.8.1　*$a\boldsymbol{A}$ 的二次风险在 $a = 1/(n+p+1)$ 时达到其极小值 $p(p+1)/(n+p+1)$. $a\boldsymbol{A}$ 的似然风险在 $a = 1/n$ (即 $a\boldsymbol{A} = \boldsymbol{S}$) 时达到极小值 $p\ln n - \sum_{i=1}^{p} \mathrm{E}\ln\chi^2_{n+1-i}$.*

证明　由损失函数的不变性知

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\Sigma}} L_q(\boldsymbol{\Sigma}, a\boldsymbol{A}) &= \mathrm{E}_{\boldsymbol{I}} L_q(\boldsymbol{I}, a\boldsymbol{A}^*) \\
&= \mathrm{E}_{\boldsymbol{I}} \operatorname{tr}(a\boldsymbol{A}^* - \boldsymbol{I})^2 \\
&= \mathrm{E}_{\boldsymbol{I}} \left(a^2 \sum_{i,j=1}^{p} a_{ij}^{*2} - 2a \sum_{i=1}^{p} a_{ii}^* + p \right) \\
&= a^2[(2n + n^2)p + np(p-1)] - 2anp + p \\
&= p[n(n+p+1)a^2 - 2na + 1],
\end{aligned} \tag{3}$$

所以上式在 $a = 1/(n+p+1)$ 时达到极小值. 类似地有

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\Sigma}} L_l(\boldsymbol{\Sigma}, a\boldsymbol{A}) &= E_{\boldsymbol{I}} L_l(\boldsymbol{I}, a\boldsymbol{A}^*) = \mathrm{E}_{\boldsymbol{I}}(a \operatorname{tr}\boldsymbol{A}^* - \ln|\boldsymbol{A}^*| - p\ln a - p) \\
&= p[na - \ln a - 1] - \mathrm{E}_{\boldsymbol{I}} \ln|\boldsymbol{A}^*|
\end{aligned} \tag{4}$$

在 $a = 1/n$ 时达到极小值.

虽然就损失函数而言, $a\boldsymbol{A}$ 的估计的极小风险是常量, 但是这个估计不是极小化极大的. 考虑满足

$$\boldsymbol{G}(\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}') = \boldsymbol{H}\boldsymbol{G}(\boldsymbol{A})\boldsymbol{H}' \tag{5}$$

的估计 $G(A)$, 其中H为下三角矩阵. 两类损失函数关于变换 $G^* = HGH'$, $\Sigma^* = H\Sigma H'$ 不变.

设 $A = I$, H 是第 i 个对角元为 -1 其余对角元为 1 的对角矩阵 D_i. 则 $HAH' = I$, (5) 的第 i,j 元满足

$$g_{ij}(I) = -g_{ij}(I), \quad j \neq i. \tag{6}$$

因此 $g_{ij}(I) = 0, j \neq i$, $G\ (I)$ 为对角矩阵, 记为 D. 因为存在下三角矩阵 T 使得 $A = TT'$, 所以

$$\begin{aligned} G(A) &= G(TIT') \\ &= TG(I)T' \\ &= TDT', \end{aligned} \tag{7}$$

其中 D 是不依赖于 A 的对角矩阵. 若对每个非奇异矩阵 H, (5) 都成立, 则存在某个 a, 使得 $D = aI$. (可以把 H 取作置换矩阵.)

由损失函数的不变性可知, 若 $\Sigma = KK'$, 其中 K 是下三角矩阵, 则

$$\begin{aligned} \mathrm{E}_{\Sigma} L[\Sigma, G(A)] &= \int L[\Sigma, G(A)] C(p,n) |\Sigma|^{-\frac{1}{2}n} |A|^{\frac{1}{2}(n-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}A} \mathrm{d}A \\ &= \int L[KK', G(A)] C(p,n) |KK'|^{-\frac{1}{2}n} |A|^{\frac{1}{2}(n-p-1)} \\ &\quad \cdot \mathrm{e}^{-\frac{1}{2}\mathrm{tr}K'^{-1}K^{-1}} \mathrm{d}A \\ &= \int L[KK', G(KA^*K')] C(p,n) |A^*|^{\frac{1}{2}(n-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}A^*} \mathrm{d}A^* \\ &= \mathrm{E}_I L[KK', KG(A^*)K'] \\ &= \mathrm{E}_I L[I, G(A^*)]. \end{aligned} \tag{8}$$

风险不依赖于 Σ.

对于二次损失函数, 计算

$$\begin{aligned} \mathrm{E}_I L_q[I, G(A)] &= \mathrm{E}_I L_q[I, TDT'] \\ &= \mathrm{E}_I \mathrm{tr}(TDT' - I)^2 \\ &= \mathrm{E}_I \mathrm{tr}(TDT'TDT' - 2TDT' + I) \\ &= \mathrm{E}_I \sum_{i,j,k,l=1}^{p} t_{ij} d_j t_{kj} t_{kl} d_l t_{il} - 2\mathrm{E}_I \sum_{i,j=1}^{p} t_{ij}^2 d_j + p. \end{aligned} \tag{9}$$

因为 t_{ii} 服从自由度为 $n+1-i$ 的 χ^2 分布, $t_{ij}\ (i > j)$ 服从 $N(0,1)$ 分布, 且 T 的 (非零) 元素是独立的, 所以计算上面的期望可得

$$\mathrm{E}_I L_q[I, G(A)] = d'Fd - 2f'd + p, \tag{10}$$

其中 $F = (f_{ij}), f = (f_i)$,

$$f_{ii}=(n+p-2i+1)(n+p-2i+3),$$
$$f_{ij}=n+p-2j+1,\quad i<j,$$
$$f_i=n+p+2i+1,\tag{11}$$

$\boldsymbol{d}=(d_1,\cdots,d_p)'$. 因为 $\boldsymbol{d}'\boldsymbol{F}\boldsymbol{d}=\mathrm{Etr}(\boldsymbol{TDT}')^2>0$, $\boldsymbol{F}$ 是正定矩阵, 而且 (10) 式在 $\boldsymbol{d}=\boldsymbol{F}^{-1}\boldsymbol{f}$ 时达到其唯一极小值 $p-\boldsymbol{f}'\boldsymbol{F}^{-1}\boldsymbol{f}$.

定理 7.8.2 对于二次损失函数, 关于线性变换 $\boldsymbol{\Sigma}\to\boldsymbol{H\Sigma H}'$, $\boldsymbol{A}\to\boldsymbol{HAH}'$ 不变的最佳估计为 $G(\boldsymbol{A})=\boldsymbol{TDT}'$, 其中 $\boldsymbol{H}$ 和 $\boldsymbol{T}$ 为下三角矩阵, $\boldsymbol{A}=\boldsymbol{TT}'$, $\boldsymbol{D}$ 是对角元为 $\boldsymbol{d}=\boldsymbol{F}^{-1}\boldsymbol{f}$ 的对角矩阵, $\boldsymbol{F}$ 和 $\boldsymbol{f}$ 由 (11) 式定义.

因为 $\boldsymbol{d}=\boldsymbol{F}^{-1}\boldsymbol{f}$ 与 $\boldsymbol{\varepsilon}=(1,\cdots,1)'$ 不成比例, 即 $\boldsymbol{F\varepsilon}$ 与 $\boldsymbol{f}$ 不成比例 (见习题 7.28), 这个估计比形如 $a\boldsymbol{A}$(它是唯一的一种关于一般线性群不变的估计) 的估计具有较小的 (二次) 损失. Kiefer (1957) 证明了如果一个估计在某估计类中是极小化极大的, 而且该估计类对满足一定条件[①] 的变换群是不变的, 那么这个估计在所有估计类中都是极小化极大的. 在这个问题中, 三角线性变换群满足这些条件, 但是不是所有的线性变换群都满足这个条件.

这个估计的定义依赖于坐标系和坐标的个数的选择.

定理 7.8.3 定理 7.8.2 中定义的估计 $G(\boldsymbol{A})$ 关于二次损失函数是极小化极大的.

当 $p=2$ 时,

$$d_1=\frac{(n+1)^2-(n-1)}{(n+1)^2(n+3)-(n-1)},\quad d_2=\frac{(n+1)(n+2)}{(n+1)^2(n+3)-(n-1)}.\tag{12}$$

风险为

$$2\frac{3n^2+5n+4}{n^3+5n^2+6n+4}.\tag{13}$$

最佳估计 $a\boldsymbol{A}$ 的风险与最佳估计 $\boldsymbol{TDT}'$ 的风险之差为

$$\frac{6}{n+3}-\frac{6n^2+10n+8}{n^3+5n^2+6n+4}=\frac{2n(n-1)}{(n+3)(n^3+5n^2+6n+4)}.\tag{14}$$

当 $n=2$ 时, 该差为 1/55 (相对于 6/5); 当 $n=3$ 时, 该差为 1/47 (相对于 1), 这是 $2/n^2$ 阶的结果; 至少对 $p=2$ 而言, 使用估计 $\boldsymbol{TDT}'$ 所带来的改进并不大.

对于似然损失函数, 我们计算

$$\begin{aligned}\mathrm{E}_{\boldsymbol{I}}L_l[\boldsymbol{I},\boldsymbol{G}(\boldsymbol{A})]&=\mathrm{E}_{\boldsymbol{I}}L_l[\boldsymbol{I},\boldsymbol{TDT}']\\&=\mathrm{E}_{\boldsymbol{I}}[\mathrm{tr}\boldsymbol{TDT}'-\ln|\boldsymbol{TDT}'|-p]\\&=\mathrm{E}_{\boldsymbol{I}}\left[\sum_{i,j=1}^{p}t_{ij}^2d_j-\sum_{i=1}^{p}\ln t_{ii}^2\right]-\sum_{i=1}^{p}\ln d_i-p\end{aligned}\tag{15}$$

① 本质条件是这个群是可分解的. 见 Kiefer (1966) 和 Kudo (1955).

$$
\begin{aligned}
&=\sum_{j=1}^{p}(n+p-2j+1)d_j-\sum_{j=1}^{p}\ln d_j\\
&\quad-\sum_{j=1}^{p}\mathrm{E}\ln\chi_{n+1-j}^2-p.
\end{aligned}
$$

当 $d_j=1/(n+p-2j+1)\,(j=1,\cdots,p)$ 时, (15) 式取得极小值.

定理 7.8.4 对于似然损失函数, 关于线性变换 $\boldsymbol{\Sigma}\to\boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}'$ 和 $\boldsymbol{A}\to\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}'$ 不变的最佳估计为 $\boldsymbol{G}(\boldsymbol{A})=\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}'$, 其中 $\boldsymbol{H}$ 和 $\boldsymbol{T}$ 为下三角矩阵, $\boldsymbol{A}=\boldsymbol{T}\boldsymbol{T}'$, $\boldsymbol{D}$ 是第 j 个对角元为 $d_j=1/(n+p-2j+1)\,(j=1,\cdots,p)$ 的对角矩阵. 极小风险为

$$
\mathrm{E}_{\boldsymbol{\Sigma}}L[\boldsymbol{\Sigma},\boldsymbol{G}(\boldsymbol{A})]=\sum_{j=1}^{p}\ln(n+p-2j+1)-\sum_{j=1}^{p}\mathrm{E}\ln\chi_{n+1-j}^2. \tag{16}
$$

定理 7.8.5 定理 7.8.4 中定义的估计 $\boldsymbol{G}(\boldsymbol{A})$ 关于似然损失函数是极小化极大的.

James and Stein (1961) 给出了这个估计. 注意到权重 $1/(n+p-1),1/(n+p-3),\cdots,1/(n-p+1)$ 的倒数关于 $1/n$ 的倒数对称.

若 $p=2$

$$
\boldsymbol{G}(\boldsymbol{A})=\frac{1}{n+1}\boldsymbol{A}+\begin{pmatrix}0&0\\0&\dfrac{2}{n^2-1}\end{pmatrix}\frac{|\boldsymbol{A}|}{a_{11}}, \tag{17}
$$

$$
\mathrm{E}[\boldsymbol{G}(\boldsymbol{A})]=\frac{n}{n+1}\boldsymbol{\Sigma}+\frac{2}{n+1}\begin{pmatrix}0&0\\0&\dfrac{|\boldsymbol{\Sigma}|}{\sigma_{11}}\end{pmatrix}. \tag{18}
$$

最佳估计 $a\boldsymbol{A}$ 的风险与最佳估计 $\boldsymbol{T}\boldsymbol{D}\boldsymbol{T}'$ 的风险之差为

$$
p\ln n-\sum_{j=1}^{p}\ln(n+p-2j+1)=-\sum_{j=1}^{p}\ln\left(1+\frac{p-2j+1}{n}\right). \tag{19}
$$

若 $p=2$, 改进量为

$$
\begin{aligned}
-\ln\left(1+\frac{1}{n}\right)-\ln\left(1-\frac{1}{n}\right)&=-\ln\left(1-\frac{1}{n^2}\right)\\
&=\frac{1}{n^2}+\frac{1}{2n^4}+\frac{1}{3n^6}+\cdots,
\end{aligned} \tag{20}
$$

当 $n=2$ 时, 该值为 0.228; 当 $n=3$ 时, 该值为 0.118; $n=4$ 时, 该值为 0.065. 风险 (19) 对任意的 p 是 $O(\frac{1}{n^2})$ 的. (见习题 7.31.)

这些估计的一个明显的缺点是他们依赖于坐标系的选择. 设第 i 个置换矩阵为 $\boldsymbol{P}_i,i=1,\cdots,p!$, 且 $\boldsymbol{P}_i\boldsymbol{A}\boldsymbol{P}_i'=\boldsymbol{T}_i\boldsymbol{T}_i'$, 其中 $\boldsymbol{T}_i$ 是下三角矩阵, 且满足 $t_{jj}>0,j=1,\cdots,p!$. 若让估计值以概率 $1/p!$ 取 $\boldsymbol{P}_i'\boldsymbol{T}_i\boldsymbol{D}\boldsymbol{T}_i'\boldsymbol{P}_i$, 则这个随机化的估计不依赖于

坐标的个数, 而且它的风险与原始坐标数的风险是相同的. 因为损失函数是凸的, 所以 $(1/p!)\sum_i \boldsymbol{P}_i'\boldsymbol{T}_i\boldsymbol{D}\boldsymbol{T}_i'\boldsymbol{P}_i$ 至少和风险函数一样好, 而此时的风险是依赖于 $\boldsymbol{\Sigma}$ 的.

Haff (1980) 证明了 $G(\boldsymbol{A}) = [1/(n+P+1)](\boldsymbol{A}+\gamma u\boldsymbol{C})$ 比 $[1/(n+p+1)]\boldsymbol{A}$ 有更小的二次风险, 其中 γ 是常数, $0 \leqslant \gamma \leqslant 2(p-1)/(n-p+3)$, $u = 1/\mathrm{tr}(\boldsymbol{A}^{-1}\boldsymbol{C})$, $\boldsymbol{C}$ 为任意正定矩阵. 估计值 $G(\boldsymbol{A}) = (1/n)[\boldsymbol{A}+ut(u)\boldsymbol{C}]$ 比 $\boldsymbol{S}$ 有更小的似然风险, 其中 $t(u)$ 是一个绝对连续且非增的函数, $0 \leqslant t(u) \leqslant 2(p-1)/n$.

7.9　椭球等高分布

7.9.1　椭球等高分布的观测

假设观测 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 来自随机向量 $\boldsymbol{X}$, $\boldsymbol{X}$ 的密度为

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g\big[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})\big]. \tag{1}$$

设 $\boldsymbol{A} = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'$, $n = N-1$, $\boldsymbol{S} = (1/n)\boldsymbol{A}$, 则当 $N\to\infty$ 时, $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Sigma}$. $\sqrt{N}\mathrm{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})$ 的极限正态分布由定理 3.6.2 给出.

7.2 节利用满足条件 $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{T}'$ 的下三角矩阵 $\boldsymbol{T}$ 得到了 $\boldsymbol{A}$ 的分布, 类似地我们可以得到$\boldsymbol{S}$的分布. 由 $\boldsymbol{S} = \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}'$ 定义下三角矩阵 $\tilde{\boldsymbol{T}}$, $\tilde{t_{ii}} \geqslant 0, i = 1,\cdots,p$. 则 $\tilde{\boldsymbol{T}} = (1/\sqrt{n})\boldsymbol{T}$. 若 $\boldsymbol{\Sigma} = \boldsymbol{I}$, 则 $\boldsymbol{S}\xrightarrow{p}\boldsymbol{I}$, $\tilde{\boldsymbol{T}}\xrightarrow{p}\boldsymbol{I}$, $\sqrt{N}(\boldsymbol{S}-\boldsymbol{I})$ 和 $\sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{I})$ 具有正态极限分布, 且

$$\sqrt{N}(\boldsymbol{S}-\boldsymbol{I}) = \sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{I}) + \sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{I})' + O_p(1). \tag{2}$$

即 $\sqrt{N}(s_{ii}-1) = 2\sqrt{N}(\tilde{t_{ii}}-1) + O_p(1)$, $\sqrt{N}s_{ij} = \sqrt{N}\tilde{t_{ij}} + O_p(1), i > j$. 当 $\boldsymbol{\Sigma} = \boldsymbol{I}$ 时, 集合 $\sqrt{N}(s_{11}-1),\cdots,\sqrt{N}(s_{pp}-1)$ 与集合 $\sqrt{N}s_{ij}\,(i > j)$ 是渐近独立的; $\sqrt{N}s_{12},\cdots,\sqrt{N}s_{p-1,p}$ 是相互渐近独立的, 每一个具有方差 $1+\kappa$; $\sqrt{N}(s_{ii}-1)$ 的极限方差为 $3\kappa+2$; $\sqrt{N}(s_{ii}-1)$ 和 $\sqrt{N}(s_{jj}-1)\,(i\neq j)$ 的极限协方差为 κ.

定理 7.9.1　*若 $\boldsymbol{\Sigma} = \boldsymbol{I}_p$, 则 $\sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{I}_p)$ 的极限分布是均值为 0 的正态分布. 对角元的方差为 $(3\kappa+2)/4$; 两个对角元的协方差为 $\kappa/4$; 非对角元的方差为 $\kappa+1$; 非对角元是互不相关的, 并且也与对角元不相关.*

设 $\boldsymbol{T} = \boldsymbol{\nu}+\boldsymbol{C}\boldsymbol{Y}$, 其中 $\boldsymbol{Y}$ 的密度为 $g(\boldsymbol{y}'\boldsymbol{y})$, $\boldsymbol{\Lambda} = \boldsymbol{C}\boldsymbol{C}'$, $\boldsymbol{\Sigma} = E(\boldsymbol{X}-\boldsymbol{\nu})(\boldsymbol{X}-\boldsymbol{\nu})' = (ER^2/p)\boldsymbol{\Lambda} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$, $\boldsymbol{C}$ 和 $\boldsymbol{\Gamma}$ 是下三角矩阵. 设 $\boldsymbol{S}$ 是来自于$\boldsymbol{X}$的 N 个样本的样本协方差阵, 而且 $\boldsymbol{S} = \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}'$. 那么 $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Sigma}$, $\tilde{\boldsymbol{T}}\xrightarrow{p}\boldsymbol{\Gamma}$, 以及

$$\sqrt{N}(\boldsymbol{S}-\boldsymbol{\Sigma}) = \sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{\Gamma})\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{\Gamma})' + O_p(1) \tag{3}$$

$\sqrt{N}(\tilde{\boldsymbol{T}}-\boldsymbol{\Gamma})$ 的极限分布是正态的, 它的协方差可以通过 (3) 式 $\sqrt{N}(\boldsymbol{S}-\boldsymbol{\Sigma})$ 的元素的协方差进行计算. 我们对 $\boldsymbol{T}$ 感兴趣, 主要是为了得到 $\boldsymbol{S}$ 的分布, 在这里不再对它进行更深的讨论.

7.9.2 椭球等高矩阵分布

基于左半球面密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$, 设 $\boldsymbol{X}$ $(N\times p)$ 的密度为

$$|\boldsymbol{C}|^{-N}g\big[\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')(\boldsymbol{C}')^{-1}\big]. \tag{4}$$

定理 7.9.2 设 $\boldsymbol{T}=(t_{ij})$ 由 $\boldsymbol{Y}'\boldsymbol{Y}=\boldsymbol{T}\boldsymbol{T}'$ $(t_{ij}=0,i<j,t_{ii}\geqslant 0)$ 定义, 若 $\boldsymbol{Y}$ 的密度为 $g(\boldsymbol{Y}'\boldsymbol{Y})$, 则 $\boldsymbol{T}$ 的密度为

$$\prod_{i=1}^{p}\left\{\frac{2\pi^{\frac{1}{2}(N+1-i)}}{\Gamma[\frac{1}{2}(N+1-i)]}t_{ii}^{N-i}\right\}g(\boldsymbol{T}\boldsymbol{T}')=\frac{2^p\pi^{\frac{1}{2}Np}}{\Gamma_p(\frac{1}{2}N)}\prod_{i=1}^{p}t_{ii}^{N-i}g(\boldsymbol{T}\boldsymbol{T}'). \tag{5}$$

证明 设 $\boldsymbol{Y}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p)$. 通过 $\boldsymbol{w}_1=\boldsymbol{v}_1,\boldsymbol{u}_1=\boldsymbol{w}_1/\|\boldsymbol{w}_1\|$, 递归定义

$$\boldsymbol{w}_i=\boldsymbol{v}_i-\sum_{j=1}^{i-1}\boldsymbol{w}_j\frac{\boldsymbol{w}_j'\boldsymbol{v}_i}{\|\boldsymbol{w}_j\|^2}=\boldsymbol{v}_i-\sum_{j=1}^{i-1}\boldsymbol{u}_j\boldsymbol{u}_j'\boldsymbol{v}_i, \tag{6}$$

其中 $\boldsymbol{u}_i=\boldsymbol{w}_i/\|\boldsymbol{w}_i\|$. 则有 $\boldsymbol{w}_i'\boldsymbol{w}_j=0$, $\boldsymbol{u}_i'\boldsymbol{u}_j=0,i\neq j$, $\boldsymbol{u}_i'\boldsymbol{u}_i=1$. 基于 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$ (即 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$), 设 $\boldsymbol{Q}_i$ 是前 $i-1$ 行为 $\boldsymbol{u}_1',\cdots,\boldsymbol{u}_{i-1}'$ 的正交矩阵; 即 (见引理 A.4.2)

$$\boldsymbol{Q}_i'=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{i-1},\boldsymbol{Q}_i^{*'}). \tag{7}$$

定义

$$z_i=\boldsymbol{Q}_i\boldsymbol{v}_i=\begin{pmatrix}t_{i1}\\ \vdots\\ t_{i,i-1}\\ \boldsymbol{z}_i^*\end{pmatrix}. \tag{8}$$

$\boldsymbol{v}_i$ 的这种变换是线性的且雅可比行列式为 1. 向量 $\boldsymbol{z}_i^*$ 有 $N+1-i$ 个分量. 注意到 $\|\boldsymbol{z}_i^*\|^2=\|\boldsymbol{w}_i\|^2$,

$$\boldsymbol{v}_i=\sum_{j=1}^{i-1}t_{ij}\boldsymbol{u}_j+\boldsymbol{w}_i=\sum_{j=1}^{i-1}t_{ij}\boldsymbol{u}_j+\boldsymbol{Q}_i^{*'}\boldsymbol{z}_i^*, \tag{9}$$

$$\boldsymbol{v}_i'\boldsymbol{v}_i=\sum_{j=1}^{i-1}t_{ij}^2+\boldsymbol{z}_i^{*'}\boldsymbol{z}_i^*=\sum_{j=1}^{i}t_{ij}^2, \tag{10}$$

$$\boldsymbol{v}_j'\boldsymbol{v}_i=\sum_{k=1}^{j}t_{jk}t_{ik},\quad j<i. \tag{11}$$

由 $\boldsymbol{Y}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_p)$ 到 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_p$ 的变换的雅可比行列式为 1.

为了得到 $\boldsymbol{T}$ 的密度, 我们把 $\boldsymbol{z}_i^*$ 变成极坐标, 并对所有的角度坐标进行积分 (见 2.7.1 节). ■

上面定理证明的方式类似于 7.2 节 (6) 的证明, 但是并没有用到正态分布的信息, 例如 $t_{ii}^2\overset{d}{=}\chi_{N+1-i}^2$. 见 Fang and Zhang (1990), 定理 3.4.1.

设 $\boldsymbol{C}$ 是满足条件 $\boldsymbol{\Lambda}=\boldsymbol{C}\boldsymbol{C}'$ 的下三角矩阵. 定义 $\boldsymbol{X}=\boldsymbol{Y}\boldsymbol{C}'$.

定理 7.9.3 若 $\boldsymbol{X}$ $(N\times p)$ 的密度为

$$|\boldsymbol{C}|^{-N}g\left[\boldsymbol{C}^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{C}')^{-1}\right]. \tag{12}$$

则满足条件 $\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{T}^*\boldsymbol{T}^{*'}$ 和 $t_{ii}\geqslant 0$ 的下三角矩阵 $\boldsymbol{T}^*$ 的密度为

$$\frac{2^p\pi^{\frac{1}{2}Np}}{\Gamma_p(\frac{1}{2}N)|\boldsymbol{\Lambda}|^{\frac{1}{2}N}}\prod_{i=1}^{p}t_{ii}^{*N-1}g\left[\boldsymbol{C}^{-1}\boldsymbol{T}^*\boldsymbol{T}^{*'}(\boldsymbol{C}')^{-1}\right]. \tag{13}$$

设 $\boldsymbol{A}=\boldsymbol{X}'\boldsymbol{X}=\boldsymbol{T}^*\boldsymbol{T}^{*'}$.

定理 7.9.4 若 $\boldsymbol{X}$ 的密度为 (12), 则 $\boldsymbol{A}=\boldsymbol{X}'\boldsymbol{X}$ 的密度为

$$\frac{\pi^{\frac{1}{2}p[N-\frac{1}{2}(p-1)]}}{\Gamma_p(\frac{1}{2}N)|\boldsymbol{\Lambda}|^{\frac{1}{2}N}}|\boldsymbol{A}|^{\frac{1}{2}(N-p+1)}g\left[\boldsymbol{C}^{-1}\boldsymbol{A}(\boldsymbol{C}')^{-1}\right]. \tag{14}$$

密度 $g(\mathrm{tr}\boldsymbol{Y}'\boldsymbol{Y})$ 的类是密度 $g(\boldsymbol{Y}'\boldsymbol{Y})$ 的一个子类. 设 $\boldsymbol{X}=\boldsymbol{\varepsilon}_N\boldsymbol{\nu}'+\boldsymbol{Y}\boldsymbol{C}'$, 则 $\boldsymbol{X}$ 的密度为

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}N}g\left[\mathrm{tr}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\boldsymbol{\nu}')'\right]. \tag{15}$$

$\boldsymbol{X}$ 的一种随机的表示为 $\boldsymbol{X}\overset{d}{=}R(\boldsymbol{C}\otimes\boldsymbol{I}_N)\mathrm{vec}\boldsymbol{U}+\boldsymbol{\nu}\otimes\boldsymbol{\varepsilon}_N$. 定理 7.9.3 和定理 7.9.4 可以被特殊化到这种形式, 故定理 3.6.5 成立.

定理 7.9.5 设 $\boldsymbol{X}$ 的密度为 (12), 其中 $\boldsymbol{\Lambda}$ 为对角矩阵. 假设 $\boldsymbol{S}=(N-1)^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')'(\boldsymbol{X}-\boldsymbol{\varepsilon}_N\bar{\boldsymbol{x}}')$ 和 $\boldsymbol{R}=(\mathrm{diag}\boldsymbol{S})^{-\frac{1}{2}}\boldsymbol{S}(\mathrm{diag}\boldsymbol{S})^{-\frac{1}{2}}$. 则 $\boldsymbol{R}$ 的密度为 7.6 节的 (9) 式.

习 题

7.1 (7.2 节) 一种把直角坐标变换为极坐标的公式为

$$\begin{aligned}
y_1&=w\sin\theta_1,\\
y_2&=w\cos\theta_1\sin\theta_2,\\
y_3&=w\cos\theta_1\cos\theta_2\sin\theta_3,\\
&\vdots\\
y_{n-1}&=w\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\sin\theta_{n-1},\\
y_n&=w\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\cos\theta_{n-1},
\end{aligned}$$

其中 $-\frac{1}{2}\pi<\theta_i\leqslant\frac{1}{2}\pi, i=1,\cdots,n-2, -\pi<\theta_{n-1}\leqslant\pi, 0\leqslant w<\infty$

(a) 证明 $w^2=\sum y_\alpha^2$. [提示: 依次计算 $y_n^2+y_{n-1}^2$, $(y_n^2+y_{n-1}^2)+y_{n-2}^2$, 等等.]

(b) 证明雅可比行列式为 $w^{n-1}\cos^{n-2}\theta_1\cos^{n-3}\theta_2\cdots\cos\theta_{n-2}$. [提示: 证明

$$\left|\frac{\partial(y_1,\cdots,y_n)}{\partial(\theta_1,\cdots,\theta_{n-1},w)}\right|$$

$$\cdot\begin{vmatrix} \cos\theta_1 & 0 & \cdots & 0 & 0 \\ 0 & \cos\theta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \cos\theta_{n-1} & 0 \\ w\sin\theta_1 & w\sin\theta_2 & \cdots & w\sin\theta_{n-1} & 1 \end{vmatrix}$$

$$= \begin{vmatrix} w & x & \cdots & x & & x & & \\ 0 & w\cos\theta_1 & \cdots & x & & x & & \\ \vdots & \vdots & & \vdots & & \vdots & & \\ 0 & 0 & \cdots & w\cos\theta_1 & \cdots \cos\theta_{n-2} & x & & \\ 0 & 0 & \cdots & 0 & & \cos\theta_1 & \cdots & \cos\theta_{n-1} \end{vmatrix},$$

其中 x 代表未知元素.]

7.2 (7.2 节) 证明

$$\int_{-\pi/2}^{\pi/2} \cos^{h-1}\theta \mathrm{d}\theta = \frac{\Gamma(\frac{1}{2}h)\Gamma(\frac{1}{2})}{\Gamma[\frac{1}{2}(h+1)]}.$$

[提示: 设 $\cos^2\theta = u$, 利用 $B(p,q)$ 的定义.]

7.3 (7.2 节) 利用习题 7.1 和习题 7.2 证明 n 维单位球面的曲面面积为

$$C(n) = \frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}.$$

7.4 (7.2 节) 利用习题 7.1~7.3 证明: 若 $\boldsymbol{y}' = (y_1, \cdots, y_n)$ 的密度为 $f(\boldsymbol{y}'\boldsymbol{y})$, 则 $u = \boldsymbol{y}'\boldsymbol{y}$ 的密度为 $\frac{1}{2}C(n)f(u)u^{\frac{1}{2}n-1}$.

7.5 (7.2 节) χ^2 分布. 利用习题 7.4 证明: 若 $y_1, \cdots, y_n$ 相互独立且服从 $N(0,1)$ 分布, 则 $U = \sum_{\alpha=1}^n y_\alpha^2$ 的密度为 $u^{\frac{1}{2}n-1}\mathrm{e}^{\frac{1}{2}u}/[2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)]$, 即自由度为 n 的 χ^2 密度.

7.6 (7.2 节) 利用 7.6 节的 (9) 式求 $\boldsymbol{A}$ 的分布.

7.7 (7.2 节) 利用定理 7.2.1 的证明说明 $\Pr\{|\boldsymbol{A}| = 0\} = 0$.

7.8 (7.2 节) 复正态分布的参数估计的独立性. 设 $z_1, \cdots, z_n$ 是来自均值为 θ 协方差阵为 $\boldsymbol{P}$ 的复正态分布的 N 个观测 (见习题 2.64). 证明 $\bar{\boldsymbol{Z}}$ 和 $\boldsymbol{A} = \sum_{\alpha=1}^N (\boldsymbol{Z}_\alpha - \bar{\boldsymbol{Z}})(\boldsymbol{Z}_\alpha - \bar{\boldsymbol{Z}})^*$ 是分布独立的, 且 $\boldsymbol{A}$ 的分布与 $\sum_{\alpha=1}^N \boldsymbol{W}_\alpha \boldsymbol{W}_\alpha^*$ 的分布相同, 其中 $\boldsymbol{W}_1, \cdots, \boldsymbol{W}_n$ 相互独立且服从均值为 $\boldsymbol{0}$ 协方差阵为 $\boldsymbol{P}$ 的复正态分布.

7.9 (7.2 节) 复 Wishart 分布. 设 $\boldsymbol{W}_1, \cdots, \boldsymbol{W}_n$ 相互独立且都服从均值为 $\boldsymbol{0}$ 协方差阵为 $\boldsymbol{P}$ 的复正态分布 (见习题 2.64). 证明 $\boldsymbol{B} = \sum_{\alpha=1}^N \boldsymbol{W}_\alpha \boldsymbol{W}_\alpha^*$ 的密度为

$$\frac{|\boldsymbol{B}|^{n-p}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{B}\boldsymbol{P}^{-1}}}{|\boldsymbol{P}|^n \pi^{\frac{1}{2}p(p-1)} \prod_{i=1}^p \Gamma(n+1-i)}.$$

7.10 (7.3 节) 利用 $W(\boldsymbol{\Sigma}, n)$ 求$\boldsymbol{A}$的特征函数. [提示: 由 $\int w(\boldsymbol{A}|\boldsymbol{\Sigma}, n)\mathrm{d}\boldsymbol{A} = 1$, 可得

$$\int \frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)} \exp(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Phi}^{-1}\boldsymbol{A})\mathrm{d}\boldsymbol{A}}{2^{\frac{1}{2}pn}\Gamma_p(\frac{1}{2}n)} = |\boldsymbol{\Phi}|^{\frac{1}{2}n}$$

是关于 $\boldsymbol{\Phi}$ 的恒等式.] 注意与 7.3.1 节 Wishart 分布的证明做对比.

7.11 (7.3.2 节) 利用特征函数证明定理 7.3.2.

7.12　(7.3.1 节)　通过对特征函数 (11) 求导的方法来求 $\boldsymbol{A}$ 的前两阶矩.

7.13　(7.3 节)　设 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_n$ 相互独立且服从 $N(\boldsymbol{0},\boldsymbol{I})$ 分布, 设 $\boldsymbol{W}=\sum_{\alpha,\beta=1}^{n} b_{\alpha\beta}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\beta'$. 证明: 若对任意满足 $\boldsymbol{a}'\boldsymbol{a}=1$ 的 $\boldsymbol{a}$ 有 $\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}=\chi_m^2$, 则 $\boldsymbol{W}$ 服从 $W(\boldsymbol{I},m)$ 分布. [提示: 利用 $\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}$ 的特征函数.]

7.14　(7.4 节)　设 $\boldsymbol{x}_\alpha\,(\alpha=1,\cdots,N)$ 是来自 $N(\boldsymbol{\beta} z_\alpha,\boldsymbol{\Sigma})$ 的观测, 其中 z_α 是标量. 设 $\boldsymbol{b}=\sum_\alpha z_\alpha\boldsymbol{x}_\alpha/\sum_\alpha z_\alpha^2$. 利用定理 7.4.1 证明 $\sum_\alpha \boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\boldsymbol{b}\boldsymbol{b}'\sum_\alpha z_\alpha^2$ 和 $\boldsymbol{b}\boldsymbol{b}'$ 是相互独立的.

7.15　(7.4 节)　利用 Γ 函数的递推公式证明

$$\mathrm{E}(\chi_{N-1}^2\chi_{N-2}^2)^h=\mathrm{E}(\chi_{2N-4}^2/4)^h,\quad h\geqslant 0;$$

χ_{N-1}^2 和 χ_{N-2}^2 是相互独立的. 并以此证明 $\chi_{N-1}^2\chi_{N-2}^2$ 的分布与 $\chi_{2N-4}^2/4$ 的分布相同.

7.16　(7.4 节)　由引理 7.4.1 证明定理 7.4.1. [提示: 证明如果 $\boldsymbol{Q}_i$ 服从分布 $W(\boldsymbol{\Sigma},r_i)$ 则 (6) 式成立, 而且由 $\boldsymbol{Q}_i$ 的独立性可知 (6) 中的$\boldsymbol{I}$没有交叠, 其中$\boldsymbol{I}$的阶为 r_i.]

7.17　(7.5 节)　由 $W(\boldsymbol{\Sigma},n)$ 直接求 $\mathrm{E}|\boldsymbol{A}|^h$. [提示: 利用

$$\int w(\boldsymbol{A}|\boldsymbol{\Sigma},n)\mathrm{d}\boldsymbol{A}\equiv 1$$

证明

$$\int|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}\exp\left(-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)\mathrm{d}\boldsymbol{A}=2^{\frac{1}{2}np}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\Gamma_p\left(\frac{1}{2}n\right)$$

关于 n 的恒等式.]

7.18　(7.5 节)　考虑由

$$N(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}^*)\leqslant\frac{(N-1)p}{N-p}F_{p,N-p}(\varepsilon)$$

确定的 $\boldsymbol{\mu}$ 的置信区域, 其中 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 是基于 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 的 N 个样本. 求置信区域的体积的期望值.

7.19　(7.6 节)　证明: 若 $\boldsymbol{\Sigma}=\boldsymbol{I}$, 则 $r_{ij\cdot p}\,(i,j=1,\cdots,p-1)$ 和 $r_{1p},\cdots,r_{p-1,p}$ 的联合密度为

$$\frac{\Gamma^{p-1}[\frac{1}{2}(n-1)]|\boldsymbol{R}_{11\cdot p}|^{\frac{1}{2}(n-p-1)}}{\pi^{(p-1)(p-2)/4}\prod_{i=1}^{p-1}\Gamma[\frac{1}{2}(n-i)]}\prod_{i=1}^{p-1}\frac{\Gamma(\frac{1}{2}n)}{\pi^{\frac{1}{2}}\Gamma[\frac{1}{2}(n-1)]}(1-r_{ip}^2)^{\frac{1}{2}(n-3)},$$

其中 $\boldsymbol{R}_{11\cdot p}=(r_{ij\cdot p})$. [提示: $r_{ij\cdot p}=(r_{ij}-r_{ip}r_{jp})/(\sqrt{1-r_{ip}^2}\sqrt{1-r_{jp}^2})$ 和 $|r_{ij}|=|\sqrt{1-r_{ip}^2}\sqrt{1-r_{jp}^2}r_{ij\cdot p}|$. 利用 (9).]

7.20　(7.6 节)　证明 $r_{12\cdot 3,\cdots,p},r_{13\cdot 4,\cdots,p},r_{23\cdot 4,\cdots,p},\cdots,r_{1p},\cdots,r_{p-1,p}$ 的联合密度为

$$\frac{\Gamma\{\frac{1}{2}[n-(p-2)]\}}{\pi^{\frac{1}{2}}\Gamma\{\frac{1}{2}[n-(p-1)]\}}(1-r_{12\cdot 3,\cdots,p}^2)^{\frac{1}{2}[n-(p+1)]}$$

$$\cdot\prod_{i=1}^{2}\frac{\Gamma\{\frac{1}{2}[n-(p-3)]\}}{\pi^{\frac{1}{2}}\Gamma\{\frac{1}{2}[n-(p-2)]\}}(1-r_{i3\cdot 4,\cdots,p}^2)^{\frac{1}{2}(n-p)}$$

$$\cdots\prod_{i=1}^{p-2}\frac{\Gamma[\frac{1}{2}(n-1)]}{\pi^{\frac{1}{2}}\Gamma[\frac{1}{2}(n-2)]}(1-r_{i,p-1\cdot p}^2)^{\frac{1}{2}(n-4)}$$

$$\cdot \prod_{i=1}^{p-1} \frac{\Gamma(\frac{1}{2}n)}{\pi^{\frac{1}{2}}\Gamma[\frac{1}{2}(n-1)]}(1-r_{ip}^2)^{\frac{1}{2}(n-3)}.$$

[提示：利用习题 7.19 的结果归纳证明.]

7.21 (7.6 节) 证明 (不能利用习题 7.20) 若 $\Sigma = I$, 则 $r_{1p},\cdots,r_{p-1,p}$ 是分布独立的. [提示: $r_{ip} = a_{ip}/(\sqrt{a_{ii}}\sqrt{a_{pp}})$. 证明当 $(z_{1p},\cdots,z_{np})$ 固定时每一对 $(a_{1p},a_{11}),\cdots,(a_{p-1,p},a_{p-1,p-1})$ 都是独立的, 而且由 4.2.1 节可知在条件 $z_{\alpha p}$ 下 r_{ip} 的边际分布不依赖于 $z_{\alpha p}$.]

7.22 (7.6 节) 证明 (不能利用习题 7.19 和习题 7.20) 若 $\Sigma = I$, 则集合 $r_{1p},\cdots,r_{p-1,p}$ 与集合 $r_{ij\cdot p}\,(i,j=1,\cdots,p-1)$ 是独立的. [提示: 由 4.3.2 节知 a_{pp} 和 (a_{ip}) 与 $(a_{ij\cdot p})$ 是独立的. 通过证明 $a_{ii\cdot p}$ 和 $(r_{ij\cdot p})$ 独立来证明 $a_{pp},(a_{ip})$ 和 $a_{ii}\,(i=1,\cdots,p-1)$ 是独立的. 见习题 4.21.]

7.23 (7.6 节) 利用习题 7.21 和习题 7.22 证明习题 7.20 的结论.

7.24 (7.6 节) 倒推习题 7.20 获得 7.6 节的 (9) 式.

7.25 (7.6 节) 证明当 $p=3$ 且 Σ 为对角矩阵时, r_{12},r_{13},r_{23} 不相互独立.

7.26 (7.6 节) 证明当 Σ 是对角矩阵时, r_{ij} 是成对独立的.

7.27 (7.7 节) *多元 t 分布*. 设 y 和 u 相互独立且分别服从 $N(0,\boldsymbol{\Sigma})$ 与 χ^2 分布, 设 $\sqrt{n/u}\,y = \boldsymbol{x}-\boldsymbol{\mu}$.

(a) 证明$\boldsymbol{x}$的密度为

$$\frac{\Gamma[\frac{1}{2}(n+p)]}{\Gamma(\frac{1}{2}n)n^{\frac{1}{2}p}\pi^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}\left[1-\frac{1}{n}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]^{\frac{1}{2}(n+p)}}.$$

(b) 证明 $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{\mu}$ 和

$$\mathrm{E}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})' = \frac{n}{n-2}\boldsymbol{\Sigma}.$$

7.28 (7.8 节) 通过计算 $\boldsymbol{F}\varepsilon$ 证明 $\boldsymbol{F}\varepsilon$ 与 $\boldsymbol{f}$ 不成比例.

7.29 (7.8 节) 证明当 $p=2$ 时

$$\boldsymbol{TDT}' = d_1\boldsymbol{A} + (d_2-d_1)\begin{pmatrix} 0 & 0 \\ 0 & \frac{|\boldsymbol{A}|}{a_{11}} \end{pmatrix}.$$

7.30 (7.8 节) 证明 (17) 和 (18). [提示: 为了证明 (18), 可以假设 $\boldsymbol{\Sigma} = \boldsymbol{KK}'$, $\boldsymbol{A} = \boldsymbol{KA}^*\boldsymbol{K}'$ 和 $\boldsymbol{A}^* = \boldsymbol{T}^*\boldsymbol{T}^*$, 其中 $\boldsymbol{K}$ 和 $\boldsymbol{T}^*$ 是下三角矩阵.]

7.31 (7.8 节) 证明对最优的 $\boldsymbol{D}$ 有

$$\begin{aligned}\mathrm{E}_I L_l(\boldsymbol{I},\boldsymbol{S}) - \mathrm{E}_I L_l(\boldsymbol{I},\boldsymbol{TDT}') &= -\sum_{i=1}^{\frac{1}{2}p}\ln\left[1-\left(\frac{p-2i+1}{n}\right)^2\right],\ p\text{ 为偶数},\\ &= -\sum_{i=1}^{\frac{1}{2}(p-1)}\ln\left[1-\left(\frac{p-2i+1}{n}\right)^2\right],\ p\text{ 为奇数}.\end{aligned}$$

7.32 (7.8 节) 证明 $L_q(\boldsymbol{\Sigma},\boldsymbol{G})$ 和 $L_l(\boldsymbol{\Sigma},\boldsymbol{G})$ 关于变换 $\boldsymbol{G}^* = \boldsymbol{CGC}'$, $\boldsymbol{\Sigma}^* = \boldsymbol{C\Sigma C}'$ 是不变的, 其中$\boldsymbol{C}$是非奇异矩阵.

7.33 (7.8 节) 证明 $L_q(\boldsymbol{\Sigma}, \boldsymbol{G})$ 是 $(\boldsymbol{g}-\boldsymbol{\sigma})'\boldsymbol{\Phi}^{-1}(\boldsymbol{g}-\boldsymbol{\sigma})$ 的倍数. 提示: 变换 $\boldsymbol{\Sigma}=\boldsymbol{I}$. 然后证明

$$\boldsymbol{\Phi}=\frac{1}{n}\begin{pmatrix} 2\boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix}$$

即可.

7.34 (7.8 节) 证明 (11).

7.35 设$\boldsymbol{Y}$的密度为 $f(y)=\begin{cases} K, \ y'y \leqslant p+2, \\ 0, \ 其他, \end{cases}$. 证明 $K=\Gamma(\frac{1}{2}p+1)/[(p+2)\pi]^{\frac{1}{2}p}$, $\mathrm{E}(\boldsymbol{Y})=0$, $\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}')=\boldsymbol{I}$.

7.36 (7.2 节) 狄利克雷分布. 设 $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_m$ 相互独立且分别服从自由度为 $p_1,\cdots,p_m$ 的 χ^2 分布. 定义 $\boldsymbol{Z}_i=\boldsymbol{Y}_i/\sum_{j=1}^m \boldsymbol{Y}_j, i=1,\cdots,m$. 证明 $\boldsymbol{Z}_1,\cdots,\boldsymbol{Z}_{m-1}$ 的密度为

$$\frac{\Gamma(\frac{1}{2}\sum_{i=1}^m p_i)}{\prod_{i=1}^m \Gamma(\frac{1}{2}p_i)} z_1^{\frac{1}{2}p_1-1}\cdots z_m^{\frac{1}{2}p_m-1}, \quad z_m \equiv 1-\sum_{i=1}^{m-1} z_i,$$

$z_i \geqslant 0, i=1,\cdots,m$.

7.37 (7.8 节) 证明若 χ^2_{N-1} 和 χ^2_{N-2} 相互独立, 则 $\chi^2_{N-1}\chi^2_{N-2}$ 的分布为 $(\chi^2_{2N-4})^2/4$. [提示: 对 $x=\chi^2_{N-1}$ 和 $y=\chi^2_{N-2}$ 的联合密度作变换 $z=2\sqrt{xy}, x=x$, 将 z 的边际密度表示为 $z^{N-3}h(z)$, 其中 $h(z)$ 是对 x 积分而得. 求 $h'(z)$, 并解微分方程. 见 Srivastava and Khatri (1979), 第 3 章.]

第 8 章　一般的线性假设检验, 多元方差分析

8.1　引　　言

在本章我们把单变量的最小二乘理论 (即回归分析) 和方差分析推广到向量变量的情况. 多元代数的理论在本质上与单变量的情形是类似的. 这就使得分布理论也类似于单变量的情况, 检验准则也类似于 F 统计量. 事实上, 给定一个单变量的检验, 我们能很快写出对应的多变量的检验. 因为基于固定效应模型的方差分析可以通过最小二乘理论得到, 所以我们可以直接得到多元的方差分析理论. 然而, 对于多元的情况, 检验的显著性有许多并行标度的选择.

对于单变量的最小二乘, 我们考虑从期望值分别为 $\boldsymbol{\beta}'\boldsymbol{z}_1,\cdots,\boldsymbol{\beta}'\boldsymbol{z}_N$ 的总体中抽取独立变量 $x_1,\cdots,x_N$, 其中 $\boldsymbol{\beta}$ 是具有 q 个分量的列向量, 每个 $\boldsymbol{z}_\alpha$ 为 q 维的已知列向量. 在总体方差相同的情况下 $\boldsymbol{\beta}'$ 的最小二乘估计为

$$\boldsymbol{b}' = \left(\sum_{\alpha=1}^{N} x_\alpha \boldsymbol{z}_\alpha'\right)\left(\sum_{\alpha=1}^{N} \boldsymbol{z}_\alpha \boldsymbol{z}_\alpha'\right)^{-1}. \tag{1}$$

若总体是正态的, 该向量也是 $\boldsymbol{\beta}$ 的极大似然估计. 公共方差 σ^2 的无偏估计为

$$s^2 = \sum_{\alpha=1}^{N}(x_\alpha - \boldsymbol{b}'\boldsymbol{z}_\alpha)^2/(N-q), \tag{2}$$

若分布是正态的, 则 σ^2 的极大似然估计为 $\hat{\sigma}^2 = (N-q)s^2/N$.

在多元的情况下 $\boldsymbol{x}_\alpha$ 是一个向量, $\boldsymbol{\beta}'$ 被矩阵 $\boldsymbol{\beta}$ 所代替, σ^2 被协方差阵 $\boldsymbol{\Sigma}$ 所代替. 在 8.2 节给出的 $\boldsymbol{\beta}$ 和 $\boldsymbol{\Sigma}$ 的估计都是矩阵, 类似于 (1) 和 (2).

我们利用 F 检验检验关于 $\boldsymbol{\beta}$ 的假设, 比如假设 $\boldsymbol{\beta} = \boldsymbol{0}$. 下面的准则等价于 F 比:

$$\frac{1}{[q/(N-q)]F+1} = \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}, \tag{3}$$

其中 $\hat{\sigma}_0^2$ 是 σ^2 在原假设下的极大似然估计. 如果要求相应于多元假设的似然比准则, 例如 $\boldsymbol{\beta} = \boldsymbol{0}$, 则只需把上式的方差用广义方差代替即可. 我们将刻画原假设下似然比准则的分布, 并求出它们矩, 还给出一些特殊的分布. 一些令人满意的近似和分位数已经列表给出 (附录 B).

假设检验问题在一些线性变换下是不变的. 其他不变准则, 包括 Lewley-Hotelling 迹, Barlett-Nanda-Pillai 迹, Roy 最大根准则, 也作了处理. 另外给出了一些功效的对比.

本章证明了 $\boldsymbol{\beta}$ 的元素的置信区域或联合置信区间可以基于似然比检验、Lewley-Hotelling 迹检验、Roy 最大根检验得到, 并具体给出了几种方差分析方法的步骤, 研究了容许性的最优性、无偏性以及所研究的功效函数的单调性. 最后, 这些理论和方法被推广到椭球等高分布.

8.2　多元线性回归中的参数估计

8.2.1　极大似然估计, 最小二乘估计

设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是 N 个独立的观测, $\boldsymbol{x}_\alpha$ 来自 $N(\boldsymbol{\beta}\boldsymbol{z}_\alpha,\boldsymbol{\Sigma})$. 一般情况下向量 $\boldsymbol{z}_\alpha$ (有 q 个分量) 是已知的, $p\times p$ 矩阵$\boldsymbol{\Sigma}$ 和 $p\times q$ 矩阵 $\boldsymbol{\beta}$ 是未知的. 我们假设 $N\geqslant p+q$, 并设

$$\boldsymbol{Z}=(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N) \tag{1}$$

的秩为 q. 利用极大似然的方法估计$\boldsymbol{\Sigma}$和 $\boldsymbol{\beta}$. 似然函数为

$$\boldsymbol{L}=(2\pi)^{-\frac{1}{2}Np}|\boldsymbol{\Sigma}^*|^{-\frac{1}{2}N}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\beta}^*\boldsymbol{z}_\alpha)'\boldsymbol{\Sigma}^{*-1}(\boldsymbol{x}_\alpha-\boldsymbol{\beta}^*\boldsymbol{z}_\alpha)\right]. \tag{2}$$

(2) 中的 $\boldsymbol{\Sigma}^*$ 和 $\boldsymbol{\beta}^*$ 是不确定的. 极大似然方法就是基于样本 $\boldsymbol{x}_1,\boldsymbol{z}_1,\cdots,\boldsymbol{x}_N,\boldsymbol{z}_N$, 给出$\boldsymbol{\Sigma}$和 $\boldsymbol{\beta}$ 的估计 $\boldsymbol{\Sigma}^*$ 和 $\boldsymbol{\beta}^*$, 使得 (2) 式达到极大. 利用下面的引理能更方便地得到相应的结论.

引理 8.2.1　设

$$\boldsymbol{B}=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{z}'_\alpha\left(\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha\boldsymbol{z}'_\alpha\right)^{-1}. \tag{3}$$

则对任意的 $p\times q$ 矩阵$\boldsymbol{F}$有

$$\begin{aligned}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{F}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{F}\boldsymbol{z}_\alpha)'=&\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)'\\&+(\boldsymbol{B}-\boldsymbol{F})\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha\boldsymbol{z}'_\alpha(\boldsymbol{B}-\boldsymbol{F})'.\end{aligned} \tag{4}$$

证明　(4) 式的左端为

$$\sum_{\alpha=1}^{N}[(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)+(\boldsymbol{B}-\boldsymbol{F})\boldsymbol{z}_\alpha][(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)+(\boldsymbol{B}-\boldsymbol{F})\boldsymbol{z}_\alpha]', \tag{5}$$

由 (3) 知,

$$\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)'=\boldsymbol{0}, \tag{6}$$

所以 (4) 式的左端等于右端. ■

$\boldsymbol{L}$ 中的指数是 $-\frac{1}{2}$ 倍的

$$\mathrm{tr}\boldsymbol{\Sigma}^{*-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{\beta}^*\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{\beta}^*\boldsymbol{z}_\alpha)'=\mathrm{tr}\boldsymbol{\Sigma}^{*-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)' \quad (7)$$
$$+\mathrm{tr}\boldsymbol{\Sigma}^{*-1}(\boldsymbol{B}-\boldsymbol{\beta}^*)\boldsymbol{A}(\boldsymbol{B}-\boldsymbol{\beta}^*)',$$

其中

$$\boldsymbol{A}=\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'. \quad (8)$$

通过极小化 (7) 式的最后一项, 似然函数关于 $\boldsymbol{\beta}^*$ 取得极大值.

引理 8.2.2 若$\boldsymbol{A}$和$\boldsymbol{G}$是正定的, 对于 $\boldsymbol{F}\neq 0$, 有 $\mathrm{tr}\boldsymbol{FAF'G}>0$.

证明 设 $\boldsymbol{A}=\boldsymbol{HH'}, \boldsymbol{G}=\boldsymbol{KK'}$. 由$\boldsymbol{H}$和$\boldsymbol{K}$是非奇异的可得 $\boldsymbol{K'FH}\neq 0$, 所以对于 $\boldsymbol{F}\neq 0$ 有

$$\mathrm{tr}\boldsymbol{FAF'G}=\mathrm{tr}\boldsymbol{FHH'F'KK'}=\mathrm{tr}\boldsymbol{K'FHH'F'K} \quad (9)$$
$$=\mathrm{tr}(\boldsymbol{K'FH})(\boldsymbol{K'FH})'>0. \quad ■$$

由 (7) 和引理知, 当 $\boldsymbol{\beta}^*=\boldsymbol{B}$ 时, $\boldsymbol{L}$ 关于 $\boldsymbol{\beta}^*$ 达到极大, 即

$$\hat{\boldsymbol{\beta}}=\boldsymbol{CA}^{-1}, \quad (10)$$

其中

$$\boldsymbol{C}=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{z}_\alpha'. \quad (11)$$

然后利用引理 3.2.2, $\boldsymbol{L}$关于 $\boldsymbol{\Sigma}^*$ 在

$$\hat{\boldsymbol{\Sigma}}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\beta}}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\hat{\boldsymbol{\beta}}\boldsymbol{z}_\alpha)' \quad (12)$$

时取得极大值. 这类似于 8.1 节 (2) 式中定义的 $\hat{\sigma}^2=(N-q)s^2/N$ 的多元的形式.

定理 8.2.1 设 $\boldsymbol{x}_\alpha$ 是来自 $N(\boldsymbol{\beta}\boldsymbol{z}_\alpha,\Sigma)$ 的一个观测, 其中 $\alpha=1,\cdots,N$, $(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N)$ 的秩为 q, 则 $\boldsymbol{\beta}$ 的极大似然估计为 (10) 式, 其中 $\boldsymbol{C}=\sum_\alpha \boldsymbol{x}_\alpha\boldsymbol{z}_\alpha'$, $\boldsymbol{A}=\sum_\alpha \boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'$. $\boldsymbol{\Sigma}$的极大似然估计为 (12) 式.

由 (12) 和 (4) 及 $\boldsymbol{F}=\mathbf{0}$, 可以得到一个有用的代数结果:

$$N\hat{\boldsymbol{\Sigma}}=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\hat{\boldsymbol{\beta}}\boldsymbol{A}\hat{\boldsymbol{\beta}}'=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\boldsymbol{CA}^{-1}\boldsymbol{C}'. \quad (13)$$

下面我们讨论该估计方法的几何解释. 设 $(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 的第 i 行是 $\boldsymbol{x}_i^*$ (有 N 个分量), 设 $(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N)$ 的第 i 行是 $\boldsymbol{z}_i^*$(有 N 个分量). 则向量 $\boldsymbol{z}_1^*,\cdots,\boldsymbol{z}_q^*$ 的线性组合 $\sum_j\hat{\beta}_{ij}\boldsymbol{z}_j^*$, 是由 $\boldsymbol{z}_1^*,\cdots,\boldsymbol{z}_q^*$ 张成的 q 维空间中的一个向量. 事实上, 在所有这些向量中, 它是最接近于 $\boldsymbol{x}_i^*$ 的, 因此它是在 q 维空间上 $\boldsymbol{x}_i^*$ 的投影. 所以向量 $\boldsymbol{x}_i^*-\sum_j\hat{\beta}_{ij}\boldsymbol{z}_j^*$ 正交于 q 维空间, 由 $\boldsymbol{x}_i^*$ 在 q 维空间上的投影指向 $\boldsymbol{x}_i^*$. 对这个向量进行变换, 使它的一个端点在原点上. 则 p 个向量 $\boldsymbol{x}_1^*-\sum_j\hat{\beta}_{1j}\boldsymbol{z}_j^*,\cdots,\boldsymbol{x}_p^*-\sum_j\hat{\beta}_{pj}\boldsymbol{z}_j^*$ 是从原点出发的向量. $N\hat{\sigma}_{ii}=(\boldsymbol{x}_i^*-\sum_j\hat{\beta}_{ij}\boldsymbol{z}_j^*)(\boldsymbol{x}_i^*-\sum_j\hat{\beta}_{ij}\boldsymbol{z}_j^*)'$ 是第 i 个向量长度

的平方, $N\hat{\sigma}_{ij} = (\boldsymbol{x}_i^* - \sum_h \hat{\beta}_{ih}\boldsymbol{z}_h^*)(\boldsymbol{x}_j^* - \sum_g \hat{\beta}_{jg}\boldsymbol{z}_g^*)'$ 是第 i 个向量的长度与第 j 个向量的长度及它们夹角余弦的乘积.

这些定义了 $\boldsymbol{\beta}$ 的极大似然估计的方程, 即 $\boldsymbol{A}\boldsymbol{B}' = \boldsymbol{C}'$, 由具有 q 个未知元的 q 个方程的 p 个集合组成. 每一个集合可以通过主元凝聚法或逐步消去法解得 (附录中的 A.5 节). 向前解法对所有 (除了最右边的) 的集合都是相同的. 利用 (13) 计算 $N\hat{\boldsymbol{\Sigma}}$, 涉及 $\hat{\boldsymbol{\beta}}\boldsymbol{A}\hat{\boldsymbol{\beta}}$ 的有效算法.

设 $\boldsymbol{X}_\alpha = (\boldsymbol{x}_{1\alpha}, \cdots, \boldsymbol{x}_{p\alpha})'$, $\boldsymbol{B} = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_p)'$ 和 $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p)'$. 则 $\mathrm{E}(\boldsymbol{x}_{i\alpha}) = \boldsymbol{\beta}_i'\boldsymbol{z}_\alpha$, $\boldsymbol{b}_i$ 是 $\boldsymbol{\beta}_i$ 的最小二乘估计. 若 $\boldsymbol{G}$ 是正定矩阵, 则 $\mathrm{tr}\boldsymbol{G}\sum_{\alpha=1}^n(\boldsymbol{x}_\alpha - \boldsymbol{F}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha - \boldsymbol{F}\boldsymbol{z}_\alpha)'$ 在 $\boldsymbol{F} = \boldsymbol{B}$ 时达到极小. 这是对$\boldsymbol{B}$为最小二乘估计的另一种理解.

8.2.2 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\boldsymbol{\Sigma}}$ 的分布

下面求 $\hat{\beta}_{ig}(i = 1, \cdots, p,\ g = 1, \cdots, q)$ 的联合分布. 因为 $\hat{\beta}_{ig}$ 是 $\boldsymbol{X}_{i\alpha}$ 的线性组合, 所以 $\hat{\beta}_{ig}$ 的联合分布是正态的. 由 (10) 我们知道

$$\begin{aligned}\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left(\sum_{\alpha=1}^N \boldsymbol{X}_\alpha \boldsymbol{z}_\alpha' \boldsymbol{A}^{-1}\right) \\ &= \sum_{\alpha=1}^N \boldsymbol{\beta}\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'\boldsymbol{A}^{-1} = \hat{\boldsymbol{\beta}}\boldsymbol{A}\boldsymbol{A}^{-1} \\ &= \boldsymbol{\beta}.\end{aligned} \tag{14}$$

所以, $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的无偏估计. $\hat{\boldsymbol{\beta}}$ 的两行 $\hat{\beta}_i'$ 与 $\hat{\beta}_j'$ 的协方差为

$$\begin{aligned}&\mathrm{E}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)' \\ &= \boldsymbol{A}^{-1}\mathrm{E}\left[\sum_{\alpha=1}^N (\boldsymbol{X}_{i\alpha} - \mathrm{E}(\boldsymbol{X}_{i\alpha}))\boldsymbol{z}_\alpha \sum_{\gamma=1}^N (\boldsymbol{X}_{j\gamma} - \mathrm{E}(\boldsymbol{X}_{j\gamma}))\right]\boldsymbol{z}_\gamma'\boldsymbol{A}^{-1} \\ &= \boldsymbol{A}^{-1}\sum_{\alpha,\gamma=1}^N \mathrm{E}\big[(\boldsymbol{X}_{i\alpha} - \mathrm{E}(\boldsymbol{X}_{i\alpha}))(\boldsymbol{X}_{j\gamma} - \mathrm{E}(\boldsymbol{X}_{j\gamma}))\big]\boldsymbol{z}_\alpha\boldsymbol{z}_\gamma'\boldsymbol{A}^{-1} \\ &= \boldsymbol{A}^{-1}\sum_{\alpha,\gamma=1}^N \delta_{\alpha\gamma}\sigma_{ij}\boldsymbol{z}_\alpha\boldsymbol{z}_\gamma'\boldsymbol{A}^{-1} \\ &= \boldsymbol{A}^{-1}\sum_{\alpha=1}^N \sigma_{ij}\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'\boldsymbol{A}^{-1} \\ &= \sigma_{ij}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{A}^{-1} \\ &= \sigma_{ij}\boldsymbol{A}^{-1}.\end{aligned} \tag{15}$$

总的来说, 拥有 pq 个分量的向量 $(\hat{\beta}_1', \cdots, \hat{\beta}_p')' = \mathrm{vec}\,\hat{\boldsymbol{\beta}}$ 服从均值为 $(\beta_1', \cdots, \beta_p')' = \mathrm{vec}\,\boldsymbol{\beta}'$ 协方差阵为

$$\begin{pmatrix} \sigma_{11}\boldsymbol{A}^{-1} & \sigma_{12}\boldsymbol{A}^{-1} & \cdots & \sigma_{1p}\boldsymbol{A}^{-1} \\ \sigma_{21}\boldsymbol{A}^{-1} & \sigma_{22}\boldsymbol{A}^{-1} & \cdots & \sigma_{2p}\boldsymbol{A}^{-1} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1}\boldsymbol{A}^{-1} & \sigma_{p2}\boldsymbol{A}^{-1} & \cdots & \sigma_{pp}\boldsymbol{A}^{-1} \end{pmatrix} \tag{16}$$

的正态分布. 矩阵 (16) 是$\boldsymbol{\Sigma}$和 $\boldsymbol{A}^{-1}$ 的克罗内克 (Kronecker) 积 (直积), 记为 $\boldsymbol{\Sigma}\otimes\boldsymbol{A}^{-1}$.

由定理 4.3.3 可得 $N\hat{\boldsymbol{\Sigma}} = \sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{x}'_\alpha - \hat{\boldsymbol{\beta}}\boldsymbol{A}\hat{\boldsymbol{\beta}}'$ 的分布为 $W(\boldsymbol{\Sigma}, N-q)$. 由此可知$\boldsymbol{\Sigma}$ 的一个无偏估计为 $\boldsymbol{S} = [N/(N-q)]\hat{\boldsymbol{\Sigma}}$.

定理 8.2.2 基于 N 个观测 (其中第 α 个观测来自 $N(\boldsymbol{\beta}\boldsymbol{z}_\alpha, \boldsymbol{\Sigma})$) 的极大似然估计 $\hat{\boldsymbol{\beta}}$ 服从正态分布, 其均值为 $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ 的第 i 和第 j 行的协方差阵为 $\sigma_{ij}\boldsymbol{A}^{-1}$, 其中 $\boldsymbol{A} = \sum_\alpha \boldsymbol{z}_\alpha\boldsymbol{z}'_\alpha$. 极大似然估计 $\hat{\boldsymbol{\Sigma}}$ 的 N 倍独立地服从 $W(\boldsymbol{\Sigma}, N-q)$ 分布, 其中 q 是 $\boldsymbol{z}_\alpha$ 的分量的个数.

则密度可以写为 [利用 (4)]

$$\frac{1}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{\frac{1}{2}N}}\exp\left(-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\left[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\boldsymbol{A}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' + N\hat{\boldsymbol{\Sigma}}\right]\right\}\right). \tag{17}$$

由此可得下面的推论.

推论 8.2.1 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\boldsymbol{\Sigma}}$ 组成了 $\boldsymbol{\beta}$ 和 $\boldsymbol{\Sigma}$ 的一个充分统计量集.

下面来看一个非常有用的定理.

定理 8.2.3 设 $\boldsymbol{X}_\alpha$ 服从 $N(\boldsymbol{\beta}\boldsymbol{z}_\alpha, \boldsymbol{\Sigma})$ 分布, $\alpha = 1,\cdots,N$, 并设 $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N$ 是独立的.

(a) 设 $\boldsymbol{w}_\alpha = \boldsymbol{H}\boldsymbol{z}_\alpha$ 和 $\boldsymbol{\Gamma} = \boldsymbol{\beta}\boldsymbol{H}^{-1}$, 则 $\boldsymbol{X}_\alpha$ 的分布为 $N(\boldsymbol{\Gamma}\boldsymbol{w}_\alpha, \boldsymbol{\Sigma})$.

(b) 若 $\boldsymbol{\Gamma}$ 是基于来自 $\boldsymbol{X}_\alpha\,(\alpha = 1,\cdots,N)$ 的观测 $\boldsymbol{x}_\alpha$, 则 $\boldsymbol{\Gamma}$ 的极大似然估计为 $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\beta}}\boldsymbol{H}^{-1}$, 其中 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的极大似然估计.

(c) $\hat{\boldsymbol{\Gamma}}(\boldsymbol{w}_\alpha\boldsymbol{w}'_\alpha)\hat{\boldsymbol{\Gamma}}' = \hat{\boldsymbol{\beta}}\boldsymbol{A}\hat{\boldsymbol{\beta}}'$, 其中 $\boldsymbol{A} = \sum_\alpha \boldsymbol{z}_\alpha\boldsymbol{z}'_\alpha$, $N\boldsymbol{\Sigma}$ 的极大似然估计为 $N\hat{\boldsymbol{\Sigma}} = \sum_\alpha \boldsymbol{x}_\alpha\boldsymbol{x}'_\alpha - \hat{\boldsymbol{\Gamma}}(\boldsymbol{w}_\alpha\boldsymbol{w}'_\alpha)\hat{\boldsymbol{\Gamma}}' = \sum_\alpha \boldsymbol{x}_\alpha\boldsymbol{x}'_\alpha - \hat{\boldsymbol{\beta}}\boldsymbol{A}\hat{\boldsymbol{\beta}}'$.

(d) $\hat{\boldsymbol{\Gamma}}$ 和 $\hat{\boldsymbol{\Sigma}}$ 相互独立.

(e) $\hat{\boldsymbol{\Gamma}}$ 服从正态分布, 其均值为$\boldsymbol{\Gamma}$, $\hat{\boldsymbol{\Gamma}}$ 的第 i 和第 j 行的协方差阵为 $\sigma_{ij}(\boldsymbol{H}\boldsymbol{A}\boldsymbol{H}')^{-1} = \sigma_{ij}\boldsymbol{H}'^{-1}\boldsymbol{A}^{-1}\boldsymbol{H}^{-1}$.

证明留给读者.

若 $\boldsymbol{F} = \sum_{\alpha=1}^{N}\boldsymbol{f}'_\alpha\boldsymbol{x}_\alpha$, 则估计 $\boldsymbol{F}$ 是 β_{ig} 的一个线性估计. 如果

$$\beta_{ig} = \mathrm{E}(\boldsymbol{F}) = \mathrm{E}\left(\sum_{\alpha=1}^{N}\boldsymbol{f}'_\alpha\boldsymbol{x}_\alpha\right) = \sum_{\alpha=1}^{N}\boldsymbol{f}'_\alpha\boldsymbol{\beta}\boldsymbol{z}_\alpha = \sum_{\alpha=1}^{N}\sum_{j=1}^{p}\sum_{h=1}^{q}f_{j\alpha}\beta_{jh}z_{h\alpha} \tag{18}$$

是关于 $\boldsymbol{\beta}$ 的一个恒等式, 即

$$\sum_{\alpha=1}^{N} f_{j\alpha} z_{h\alpha} = \begin{cases} 1, & j=i,\ h=g, \\ 0, & \text{其他}, \end{cases} \tag{19}$$

则 $\boldsymbol{F}$ 是 β_{ig} 的一个线性无偏估计. 称一个线性无偏估计为最佳的, 若它在所有线性无偏估计中具有最小方差. 即对于任意的 $\boldsymbol{G} = \sum_{\alpha=1}^{N} \boldsymbol{g}'_{\alpha} \boldsymbol{x}_{\alpha}$ 和 $\mathrm{E}(\boldsymbol{G}) = \beta_{ig}$, 有 $\mathrm{E}(\boldsymbol{F} - \beta_{ig})^2 \leqslant \mathrm{E}(\boldsymbol{G} - \beta_{ig})^2$.

定理 8.2.4　β_{ig} 的最小二乘估计是它的最佳线性无偏估计.

证明　设 β_{ig} 的任意一个线性无偏估计为 $\tilde{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} f_{j\alpha} x_{j\alpha}$, 并设 $\hat{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{h=1}^{q} x_{i\alpha} z_{h\alpha} a^{hg}$ 为 β_{ig} 的最小二乘估计, 其中 $\boldsymbol{A} = \sum_{\alpha=1}^{N} \boldsymbol{z}_{\alpha} \boldsymbol{z}'_{\alpha}$. 则

$$\begin{aligned} \mathrm{E}(\tilde{\beta}_{ig} - \beta_{ig})^2 &= \mathrm{E}[\hat{\beta}_{ig} - \beta_{ig} + (\tilde{\beta}_{ig} - \hat{\beta}_{ig})]^2 \\ &= \mathrm{E}(\hat{\beta}_{ig} - \beta_{ig})^2 + 2\mathrm{E}(\hat{\beta}_{ig} - \beta_{ig})(\tilde{\beta}_{ig} - \hat{\beta}_{ig}) + \mathrm{E}(\tilde{\beta}_{ig} - \hat{\beta}_{ig})^2. \end{aligned} \tag{20}$$

由于 $\tilde{\beta}_{ig}$ 和 $\hat{\beta}_{ig}$ 是无偏的, 所以 $\tilde{\beta}_{ig} - \beta_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} f_{j\alpha} u_{j\alpha}$, $\hat{\beta}_{ig} - \beta_{ig} = \sum_{\alpha=1}^{N} \sum_{h=1}^{q} u_{i\alpha} z_{h\alpha} a^{hg}$,

$$\tilde{\beta}_{ig} - \hat{\beta}_{ig} = \sum_{\alpha=1}^{N} \sum_{j=1}^{p} \left(f_{j\alpha}^{ig} - \delta_{ij} \sum_{h=1}^{q} z_{h\alpha} a^{hg} \right) u_{j\alpha}, \tag{21}$$

其中 $\delta_{ii} = 1$ 和 $\delta_{ij} = 0, i \neq j$. 则

$$\begin{aligned} &\mathrm{E}(\hat{\beta}_{ig} - \beta_{ig})(\tilde{\beta}_{ig} - \hat{\beta}_{ig}) \\ =&\mathrm{E} \sum_{\alpha,\gamma=1}^{N} \sum_{h=1}^{q} z_{h\alpha} a^{hg} u_{i\alpha} \sum_{j=1}^{p} \left(f_{j\gamma} - \delta_{ij} \sum_{h'=1}^{q} z_{h'\gamma} a^{h'g} \right) u_{j\gamma} \\ =&\sum_{\alpha=1}^{N} \sum_{h=1}^{q} \sum_{j=1}^{p} z_{h\alpha} a^{hg} \left(f_{j\alpha} - \delta_{ij} \sum_{h'=1}^{q} z_{h'\alpha} a^{h'g} \right) \sigma_{ij} \\ =&\sigma_{ii} a^{gg} - \sigma_{ii} \sum_{h=1}^{q} \sum_{h'=1}^{q} a_{hh'} a^{hg} a^{h'g} \\ =&0. \end{aligned} \tag{22, 23}$$

由 (20) 式知 $\mathrm{E}(\tilde{\beta}_{ig} - \beta_{ig})^2 \geqslant \mathrm{E}(\hat{\beta}_{ig} - \beta_{ig})^2$. ■

8.3　关于回归系数线性假设检验的似然比准则

8.3.1　似然比准则

首先进行分块

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \tag{1}$$

其中 $\boldsymbol{\beta}_1$ 有 q_1 列, $\boldsymbol{\beta}_2$ 有 q_2 列. 得到检验线性假设

$$H:\ \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^* \tag{2}$$

的似然比准则, 其中 $\boldsymbol{\beta}_1^*$ 是给定的矩阵. 样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的极大似然函数 L 为

$$\max_{\boldsymbol{\beta},\boldsymbol{\Sigma}} L=(2\pi)^{-\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\Omega|^{-\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}pN}, \tag{3}$$

其中 $\hat{\boldsymbol{\Sigma}}_\Omega$ 由 8.2 节中的 (12) 或 (13) 给出.

为了找到由 (2) 式定义的把参数限制在 ω 上的极大似然函数, 我们设

$$\boldsymbol{y}_\alpha=\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)},\quad \alpha=1,\cdots,N, \tag{4}$$

其中

$$\boldsymbol{z}_\alpha=\begin{pmatrix}\boldsymbol{z}_\alpha^{(1)}\\ \boldsymbol{z}_\alpha^{(2)}\end{pmatrix},\quad \alpha=1,\cdots,N \tag{5}$$

的分块方式与 $\boldsymbol{\beta}$ 的相同. 则 $\boldsymbol{y}_\alpha$ 可以看成是来自 $N(\boldsymbol{\beta}_2\boldsymbol{z}_\alpha^{(2)},\boldsymbol{\Sigma})$ 的一个观测. 利用 8.2 节的方法得到 $\boldsymbol{\beta}_2$ 的估计为

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{2\omega}&=\sum_{\alpha=1}^N\boldsymbol{y}_\alpha\boldsymbol{z}_\alpha^{(2)\prime}\boldsymbol{A}_{22}^{-1}=\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)})\boldsymbol{z}_\alpha^{(2)\prime}\boldsymbol{A}_{22}^{-1}\\&=(\boldsymbol{C}_2-\boldsymbol{\beta}_1^*\boldsymbol{A}_{12})\boldsymbol{A}_{22}^{-1},\end{aligned} \tag{6}$$

其中 $\boldsymbol{C}$ 和 $\boldsymbol{A}$ 的分块方式与 $\boldsymbol{\beta}$ 和 $\boldsymbol{z}_\alpha$ 的分块方式相同,

$$\boldsymbol{C}=(\boldsymbol{C}_1\quad \boldsymbol{C}_2), \tag{7}$$

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\ \boldsymbol{A}_{21}&\boldsymbol{A}_{22}\end{pmatrix}. \tag{8}$$

$\boldsymbol{\Sigma}$ 的估计由下式给出:

$$\begin{aligned}N\hat{\boldsymbol{\Sigma}}_\omega&=\sum_{\alpha=1}^N(\boldsymbol{y}_\alpha-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{z}_\alpha^{(2)})(\boldsymbol{y}_\alpha-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{z}_\alpha^{(2)})'\\&=\sum_{\alpha=1}^N\boldsymbol{y}_\alpha\boldsymbol{y}_\alpha'-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'\\&=\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)})(\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)})'-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'.\end{aligned} \tag{9}$$

因此在 ω 上的极大似然函数为

$$\max_{\boldsymbol{\beta}_2,\boldsymbol{\Sigma}} L=(2\pi)^{-\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\omega|^{-\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}pN}. \tag{10}$$

从而检验 H 的似然比准则为 (3) 式除以 (10) 式, 即

$$\lambda=\frac{|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}}{|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}. \tag{11}$$

在检验 H 时, 当 $\lambda<\lambda_0$ 时, 我们拒绝假设, 其中 λ_0 是被选定的一个适当的数.

Hotelling T^2 准则是这个问题的特殊情况. 若 $q = q_1 = 1\,(q_2 = 0), \boldsymbol{z}_\alpha = 1, \alpha = 1, \cdots, N$, $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\mu}$, 则检验假设 $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ 的 T^2 准则是 $\boldsymbol{\beta}_1^* = \boldsymbol{\mu}_0$ 下的 (11) 式的单调函数.

对于非奇异矩阵 $\boldsymbol{D}$, 假设 $\boldsymbol{\mu} = \mathbf{0}$ 和 T^2 统计量是关于变换 $\boldsymbol{X}^* = \boldsymbol{D}\boldsymbol{X}$ 和 $\boldsymbol{x}_\alpha^* = \boldsymbol{D}\boldsymbol{x}_\alpha\,(\alpha = 1, \cdots, N)$ 不变的. 类似地, 本问题中的原假设 $\boldsymbol{\beta}_1 = \mathbf{0}$ 和检验它的似然比准则关于非奇异的线性变换也是不变的.

定理 8.3.1 对非奇异矩阵 $\boldsymbol{D}$, 检验原假设 $\boldsymbol{\beta}_1 = \mathbf{0}$ 的似然比准则 (11) 关于变换 $\boldsymbol{x}_\alpha^* = \boldsymbol{D}\boldsymbol{x}_\alpha\,(\alpha = 1, \cdots, N)$ 是不变的.

证明 关于 $\boldsymbol{x}_\alpha^*$ 的估计为

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{D}\boldsymbol{C}\boldsymbol{A}^{-1} = \boldsymbol{D}\hat{\boldsymbol{\beta}}, \tag{12}$$

$$\hat{\boldsymbol{\Sigma}}_\Omega^* = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{D}\boldsymbol{x}_\alpha - \boldsymbol{D}\hat{\boldsymbol{\beta}}\boldsymbol{z}_\alpha)(\boldsymbol{D}\boldsymbol{x}_\alpha - \boldsymbol{D}\hat{\boldsymbol{\beta}}\boldsymbol{z}_\alpha)' = \boldsymbol{D}\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{D}', \tag{13}$$

$$\hat{\boldsymbol{\beta}}_{2\omega}^* = \boldsymbol{D}\boldsymbol{C}_2\boldsymbol{A}_{22}^{-1} = \boldsymbol{D}\hat{\boldsymbol{\beta}}_{2\omega}, \tag{14}$$

$$\hat{\boldsymbol{\Sigma}}_\omega^* = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{D}\boldsymbol{x}_\alpha - \boldsymbol{D}\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{z}_\alpha^{(2)})(\boldsymbol{D}\boldsymbol{x}_\alpha - \boldsymbol{D}\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{z}_\alpha^{(2)})' = \boldsymbol{D}\hat{\boldsymbol{\Sigma}}_\omega\boldsymbol{D}'. \quad \blacksquare \tag{15}$$

8.3.2 几何解释

对于这里的代数知识, 我们可以给出其几何解释. 应用下面的引理会更方便些.

引理 8.3.1

$$\hat{\boldsymbol{\beta}}_{2\omega} - \hat{\boldsymbol{\beta}}_{2\Omega} = (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}. \tag{16}$$

证明 把正规方程 $\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{A} = \boldsymbol{C}$ 写成分块的形式

$$(\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{A}_{11} + \hat{\boldsymbol{\beta}}_{2\Omega}\boldsymbol{A}_{21}, \hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{A}_{12} + \hat{\boldsymbol{\beta}}_{2\Omega}\boldsymbol{A}_{22}) = (\boldsymbol{C}_1, \boldsymbol{C}_2). \tag{17}$$

则 $\hat{\boldsymbol{\beta}}_{2\Omega} = \boldsymbol{C}_2\boldsymbol{A}_{22}^{-1} - \hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}$. 对照 (6) 式可知引理成立.

我们可以把

$$\begin{aligned}\boldsymbol{X} - \boldsymbol{\beta}\boldsymbol{Z} &= (\boldsymbol{X} - \hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z}) + (\hat{\boldsymbol{\beta}}_{2\Omega} - \boldsymbol{\beta}_2)\boldsymbol{Z}_2 + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{Z}_1 \\ &= (\boldsymbol{X} - \hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z}) + (\hat{\boldsymbol{\beta}}_{2\omega} - \boldsymbol{\beta}_2)\boldsymbol{Z}_2 \\ &\quad -(\hat{\boldsymbol{\beta}}_{2\omega} - \hat{\boldsymbol{\beta}}_{2\Omega})\boldsymbol{Z}_2 + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{Z}_1 \\ &= (\boldsymbol{X} - \hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z}) + (\hat{\boldsymbol{\beta}}_{2\omega} - \boldsymbol{\beta}_2)\boldsymbol{Z}_2 \\ &\quad +(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)(\boldsymbol{Z}_1 - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)\end{aligned} \tag{18}$$

写成等式, 其中 $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N), \boldsymbol{Z}_1 = (\boldsymbol{z}_1^{(1)}, \cdots, \boldsymbol{z}_N^{(1)})$ 和 $\boldsymbol{Z}_2 = (\boldsymbol{z}_1^{(2)}, \cdots, \boldsymbol{z}_N^{(2)})$. $\boldsymbol{Z} = (\boldsymbol{z}_1', \boldsymbol{z}_2')$ 的行张成了 N 维空间的一个 q 维子空间. $\boldsymbol{\beta}\boldsymbol{Z}$ 的每一行都是 q 维空间中的向量, 因此 $\boldsymbol{X} - \boldsymbol{\beta}\boldsymbol{Z}$ 的每一行都是相应于 $\boldsymbol{X}$ 的行向量的 q 维空间中的向量.

在上面, $\boldsymbol{X}-\boldsymbol{\beta Z}$ 的每一个行向量都被表示成三个行向量的和. (18) 式右端的第一个矩阵的第 i 行与 q 维子空间正交, 这就导致了 $\boldsymbol{X}$ 的第 i 行向量也与 q 维子空间正交 (将在下面进行说明). $(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)\boldsymbol{Z}_2$ 的行向量空间是由 $\boldsymbol{Z}_2$ 的行向量张成的 q_2 维空间 (因为它们是 $\boldsymbol{Z}_2$ 的行向量的线性组合), $(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)$ 的行向量包含在 $\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2$ 的 q_1 维空间中, 该空间又包含在 $\boldsymbol{Z}$ 的 q 维空间中, 并且正交于 $\boldsymbol{Z}_2$ 的 q_2 维空间 [因为 $(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)\boldsymbol{Z}_2'=0$]. 因此, 见图 8.1, $\boldsymbol{X}-\boldsymbol{\beta Z}$ 的每一行都可以表示成三个正交向量的和: 一个在正交于 $\boldsymbol{Z}$ 的空间内, 一个在 $\boldsymbol{Z}_2$ 的空间内, 另一个在 $\boldsymbol{Z}$ 的子空间内且正交于 $\boldsymbol{Z}_2$.

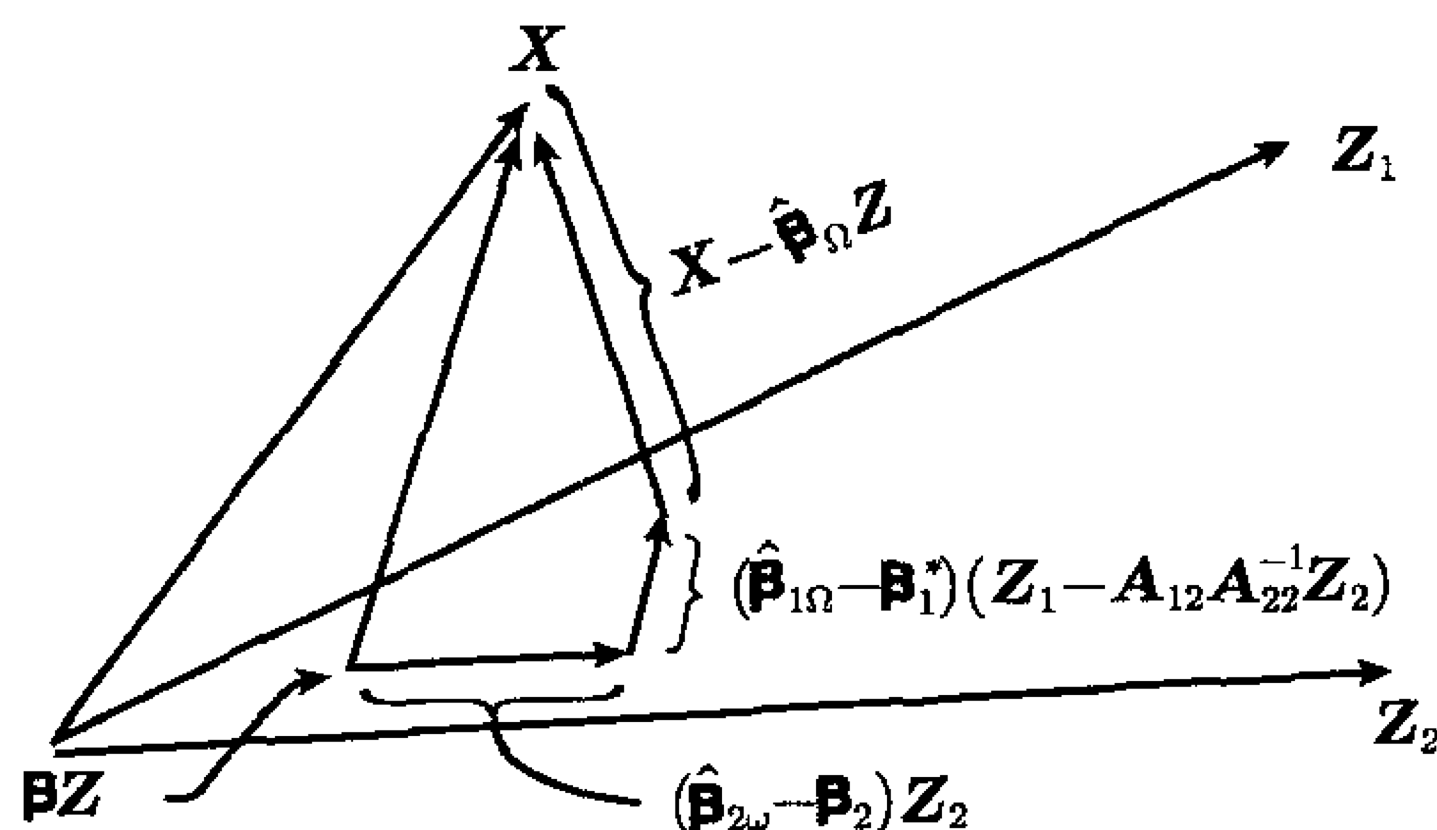


图 8.1

由正交关系可得

$$\begin{aligned}
&(\boldsymbol{X}-\boldsymbol{\beta Z})(\boldsymbol{X}-\boldsymbol{\beta Z})' \\
=&(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})'+(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)\boldsymbol{Z}_2\boldsymbol{Z}_2'(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \\
&+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)'(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)' \\
=&N\hat{\boldsymbol{\Sigma}}_\Omega+(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)\boldsymbol{A}_{22}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \\
&+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'.
\end{aligned} \tag{19}$$

(18) 式的两端同时减去 $(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)\boldsymbol{Z}_2$ 可得

$$\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2=(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2). \tag{20}$$

所以

$$\begin{aligned}
N\hat{\boldsymbol{\Sigma}}_\omega=&(\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2)(\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2)' \\
=&(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})' \\
&+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)'(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)' \\
=&N\hat{\boldsymbol{\Sigma}}_\Omega+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'.
\end{aligned} \tag{21}$$

行列式 $|\hat{\boldsymbol{\Sigma}}_\Omega|=(1/N^p)|(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})(\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z})'|$ 与 $\boldsymbol{X}-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{Z}$ 的行向量张成的超平行体的体积的平方成比例 (变换到原点). 行列式 $|\hat{\boldsymbol{\Sigma}}_\omega|=(1/N^p)|(\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-$

$\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2)(\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2)'|$ 与 $\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{Z}_2$ 的行向量张成的超平行体的体积的平方成比例 (变换到原点); 这些向量的每一个都是 $\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1$ 中正交于 $\boldsymbol{Z}_2$ 的一部分. 因此基于似然比准则的检验依赖于超平行体的体积之比. 一个是包含正交于$\boldsymbol{Z}$的向量的超平行体, 一个是包含正交于 $\boldsymbol{Z}_2$ 的向量的超平行体.

由 (15) 知, $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 的密度可以写为

$$\frac{1}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{\frac{1}{2}N}}\exp\Bigg(-\frac{1}{2}\mathrm{tr}\Bigg\{\boldsymbol{\Sigma}^{-1}\Bigg[N\hat{\boldsymbol{\Sigma}}+(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)\boldsymbol{A}_{22}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)' \tag{22}$$

$$+(\hat{\boldsymbol{\beta}}_{1\omega}-\boldsymbol{\beta}_1^*)(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'\Bigg]\Bigg\}\Bigg).$$

因此, $\hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\beta}}_{1\Omega}$ 和 $\hat{\boldsymbol{\beta}}_{2\omega}$ 形成了 $\boldsymbol{\Sigma}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ 的一个充分统计量集.

Wilk(1932) 首先给出了检验来自不同总体均值向量相等的似然比准则 (8.8 节). Wilk(1934) 和 Bartlett(1934) 将其应用推广到回归系数.

8.3.3 典范型

如果把观测的分布写成典范型将给我们研究准则的分布带来方便. 这就相当于在 N 维空间中选择一个坐标系, 使得前 q_1 个坐标轴取自 $\boldsymbol{Z}$ 的空间且正交于 $\boldsymbol{Z}_2$, 接着的 q_2 个坐标轴取自 $\boldsymbol{Z}_2$ 的空间, 剩下的 $n(=N-q)$ 个坐标轴正交于 $\boldsymbol{Z}$ 的空间.

设 $\boldsymbol{P}_2$ 为 $q_2\times q_2$ 矩阵, 且满足

$$\boldsymbol{I}=\boldsymbol{P}_2\boldsymbol{A}_{22}\boldsymbol{P}_2'=(\boldsymbol{P}_2\boldsymbol{Z}_2)(\boldsymbol{P}_2\boldsymbol{Z}_2)', \tag{23}$$

设 $\boldsymbol{P}_1$ 为 $q_1\times q_1$ 矩阵, 且满足 $(\boldsymbol{A}_{11\cdot2}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})$, 则

$$\boldsymbol{I}=\boldsymbol{P}_1\boldsymbol{A}_{11\cdot2}\boldsymbol{P}_1'=[\boldsymbol{P}_1(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)][\boldsymbol{P}_1(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)]'. \tag{24}$$

我们定义 $N\times N$ 的正交矩阵 $\boldsymbol{Q}$ 为

$$\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{Q}_1\\ \boldsymbol{Q}_2\\ \boldsymbol{Q}_3\end{pmatrix}=\begin{pmatrix}\boldsymbol{P}_1(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)\\ \boldsymbol{P}_2\boldsymbol{Z}_2\\ \boldsymbol{Q}_3\end{pmatrix}, \tag{25}$$

其中 $\boldsymbol{Q}_3$ 是任一个能使 $\boldsymbol{Q}$ 为正交矩阵的 $n\times N$ 矩阵. 那么

$$\boldsymbol{W}=(\boldsymbol{W}_1\quad \boldsymbol{W}_2\quad \boldsymbol{W}_3)=\boldsymbol{X}\boldsymbol{Q}'=\boldsymbol{X}(\boldsymbol{Q}_1'\quad \boldsymbol{Q}_2'\quad \boldsymbol{Q}_3') \tag{26}$$

的列相互独立地服从协方差阵为 $\boldsymbol{\Sigma}$ 的正态分布 (定理 3.3.1). 所以

$$\mathrm{E}(\boldsymbol{W}_1)=\mathrm{E}(\boldsymbol{X}\boldsymbol{Q}_1')=(\boldsymbol{\beta}_1\boldsymbol{Z}_1+\boldsymbol{\beta}_2\boldsymbol{Z}_2)(\boldsymbol{Z}_1-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{Z}_2)'\boldsymbol{P}_1' \tag{27}$$

$$=\boldsymbol{\beta}_1\boldsymbol{A}_{11\cdot2}\boldsymbol{P}_1'=\boldsymbol{\beta}_1\boldsymbol{P}_1^{-1},$$

$$\mathrm{E}(\boldsymbol{W}_2)=\mathrm{E}(\boldsymbol{X}\boldsymbol{Q}_2')=(\boldsymbol{\beta}_1\boldsymbol{Z}_1+\boldsymbol{\beta}_2\boldsymbol{Z}_2)\boldsymbol{Z}_2'\boldsymbol{P}_2', \tag{28}$$

$$=(\boldsymbol{\beta}_1\boldsymbol{A}_{12}+\boldsymbol{\beta}_2\boldsymbol{A}_{22})\boldsymbol{P}_2',$$

$$\mathrm{E}(\boldsymbol{W}_3)=\mathrm{E}(\boldsymbol{X}\boldsymbol{Q}_3')=\boldsymbol{\beta}\boldsymbol{Z}\boldsymbol{Q}_3'=0. \tag{29}$$

设

$$\boldsymbol{\Gamma}_1=(\gamma_1,\cdots,\gamma_{q_1})=\boldsymbol{\beta}_1\boldsymbol{A}_{11\cdot2}\boldsymbol{P}_1'=\boldsymbol{\beta}_1\boldsymbol{P}_1^{-1}, \tag{30}$$

$$\boldsymbol{\Gamma}_2=(\gamma_{q_1+1},\cdots,\gamma_q)=(\boldsymbol{\beta}_1\boldsymbol{A}_{12}+\boldsymbol{\beta}_2\boldsymbol{A}_{22})\boldsymbol{P}_2', \tag{31}$$

$$\boldsymbol{W}=(\boldsymbol{W}_1\quad\boldsymbol{W}_2\quad\boldsymbol{W}_3)=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{q_1},\boldsymbol{w}_{q_1+1},\cdots,\boldsymbol{w}_q,\boldsymbol{w}_{q+1},\cdots,\boldsymbol{w}_N). \tag{32}$$

则 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_N$ 相互独立地服从协方差阵为 $\boldsymbol{\Sigma}$ 的正态分布, 且 $\mathrm{E}(\boldsymbol{w}_\alpha)=\gamma_\alpha,\alpha=1,\cdots,q$, $\mathrm{E}(\boldsymbol{w}_\alpha)=\boldsymbol{0},\alpha=q+1,\cdots,N$.

假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 可以转换为 $\boldsymbol{\beta}_1=\boldsymbol{0}$, 即类似于 8.3.1 节, 令 $\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)}=\boldsymbol{y}_\alpha$. 在典范型下, 假设变为 $\boldsymbol{\Gamma}_1=0$. 我们可以在典范型下研究该问题, 如果需要可以再把解变回关于 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 的形式.

在 $\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{A}=\boldsymbol{C}$ 的分块形式 (17) 式中, 消去 $\hat{\boldsymbol{\beta}}_{2\Omega}$ 可得

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{1\Omega}(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})&=\boldsymbol{C}_1-\boldsymbol{C}_2\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\\&=\boldsymbol{X}(\boldsymbol{Z}_1'-\boldsymbol{Z}_2'\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\\&=\boldsymbol{W}_1\boldsymbol{P}_1'^{-1},\end{aligned} \tag{33}$$

即 $\boldsymbol{W}_1=\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{A}_{11\cdot2}\boldsymbol{P}_1'=\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{P}_1^{-1}$ 和 $\boldsymbol{\Gamma}_1=\boldsymbol{\beta}_1\boldsymbol{P}_1^{-1}$. 类似地, 由 (6) 式可得

$$\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}+\boldsymbol{\beta}_1^*\boldsymbol{A}_{12}=\boldsymbol{C}_2=\boldsymbol{X}\boldsymbol{Z}_2'=\boldsymbol{W}_2\boldsymbol{P}_2'^{-1}, \tag{34}$$

即 $\boldsymbol{W}_2=(\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}+\boldsymbol{\beta}_1^*\boldsymbol{A}_{12})\boldsymbol{P}_2'=\boldsymbol{\beta}_{2\omega}\boldsymbol{P}_2^{-1}+\boldsymbol{\beta}_1^*\boldsymbol{A}_{12}\boldsymbol{P}_2^{-1}$ 和 $\boldsymbol{\Gamma}_2=\boldsymbol{\beta}_2\boldsymbol{P}_2^{-1}+\boldsymbol{\beta}_1\boldsymbol{A}_{12}\boldsymbol{P}_2^{-1}$.

8.4 假设成立时似然比准则的分布

8.4.1 分布的特征

似然比准则是

$$U=\lambda^{2/N}=\frac{|\hat{\boldsymbol{\Sigma}}_\Omega|}{|\hat{\boldsymbol{\Sigma}}_\omega|}=\frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\Omega+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'|} \tag{1}$$

的 $\frac{1}{2}N$ 次方, 其中 $\boldsymbol{A}_{11\cdot2}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$. 下面研究当 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 时 U 的分布和它的矩. 在 8.2 节我们知道 $N\hat{\boldsymbol{\Sigma}}_\Omega$ 的分布为 $W(\boldsymbol{\Sigma},n)$, 其中 $n=N-q$, $\hat{\boldsymbol{\beta}}_\Omega-\boldsymbol{\beta}$ 的元素的联合分布为正态的且与 $N\hat{\boldsymbol{\Sigma}}_\Omega$ 独立.

由 8.3 节的 (33) 和 (24) 可得

$$\begin{aligned}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'&=(\boldsymbol{W}_1-\boldsymbol{\Gamma}_1)\boldsymbol{P}_1\boldsymbol{A}_{11\cdot2}\boldsymbol{P}_1'(\boldsymbol{W}_1-\boldsymbol{\Gamma}_1)'\\&=(\boldsymbol{W}_1-\boldsymbol{\Gamma}_1)(\boldsymbol{W}_1-\boldsymbol{\Gamma}_1)',\end{aligned} \tag{2}$$

$(\boldsymbol{W}_1 - \boldsymbol{\Gamma}_1)$ 的列相互独立且都服从 $N(0, \boldsymbol{\Sigma})$ 分布.

引理 8.4.1 $(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)'$ 服从 $W(\boldsymbol{\Sigma}, q_1)$ 分布.

引理 8.4.2 准则 U 服从分布

$$U = \frac{|\boldsymbol{G}|}{|\boldsymbol{G}+\boldsymbol{H}|}, \tag{3}$$

其中 $\boldsymbol{G}$ 服从 $W(\boldsymbol{\Sigma}, n)$ 分布, $\boldsymbol{H}$ 服从 $W(\boldsymbol{\Sigma}, m)\,(m = q_1)$ 分布, 且 $\boldsymbol{G}$ 和 $\boldsymbol{H}$ 相互独立.

设

$$\boldsymbol{G} = N\hat{\boldsymbol{\Sigma}}_\Omega = \boldsymbol{X}\boldsymbol{X}' - \boldsymbol{X}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z}\boldsymbol{X}', \tag{4}$$

$$\begin{aligned}\boldsymbol{G}+\boldsymbol{H} &= N\hat{\boldsymbol{\Sigma}}_\Omega + (\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)' \\ &= N\hat{\boldsymbol{\Sigma}}_\omega = \boldsymbol{Y}\boldsymbol{Y}' - \boldsymbol{Y}\boldsymbol{Z}_2'(\boldsymbol{Z}_2\boldsymbol{Z}_2')^{-1}\boldsymbol{Z}_2\boldsymbol{Y}',\end{aligned} \tag{5}$$

其中 $\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{\beta}_1^*\boldsymbol{Z}_1 = \boldsymbol{X} - (\boldsymbol{\beta}_1^* \quad 0)\boldsymbol{Z}$. 则

$$\boldsymbol{G} = \boldsymbol{Y}\boldsymbol{Y}' - \boldsymbol{Y}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z}\boldsymbol{Y}'. \tag{6}$$

我们把该准则记为 $U_{p,m,n}$, 其中 p 代表维数, $m = q_1$ 为 $\boldsymbol{\beta}_1$ 的列数, $n = N - q$ 是 $\boldsymbol{G}$ 的自由度.

下面我们将刻画作为 β 变量 (5.2 节) 乘积的 U 的分布. 把准则 U 表示为

$$U = V_1 V_2 \cdots V_p, \tag{7}$$

其中 $V_1 = g_{11}/(g_{11} + h_{11})$,

$$V_i = \frac{|\boldsymbol{G}_i|}{|\boldsymbol{G}_{i-1}|}\bigg/\frac{|\boldsymbol{G}_i + \boldsymbol{H}_i|}{|\boldsymbol{G}_{i-1} + \boldsymbol{H}_{i-1}|}, \quad i = 2, \cdots, p, \tag{8}$$

$\boldsymbol{G}_i$ 和 $\boldsymbol{H}_i$ 分别为 $\boldsymbol{G}$ 和 $\boldsymbol{H}$ 的前 i 行和 i 列的子矩阵. 相应地, 设 $\boldsymbol{y}_\alpha^{(i)}$ 为 $\boldsymbol{y}_\alpha = \boldsymbol{x}_\alpha - \boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)}\,(\alpha = 1, \cdots, N)$ 的前 i 个分量. 我们将证明, $\boldsymbol{V}_i$ 为 $\boldsymbol{y}_i^* = (y_{i1}, \cdots, y_{iN})$ 与它在 $\boldsymbol{Z}$ 中的投影向量的距离的平方, $\boldsymbol{Y}_{i-1} = (\boldsymbol{y}_1^{(i-1)}, \cdots, \boldsymbol{y}_N^{(i-1)})$ 可以被 $\boldsymbol{y}_i^*$ 与它在 $\boldsymbol{Z}_2$ 中投影的距离的平方整除.

引理 8.4.3 设 $\boldsymbol{y}$ 为具有 N 个分量的行向量, $\boldsymbol{U}$为 $r \times N$ 矩阵. $\boldsymbol{y}$ 关于 $\boldsymbol{U}$ 回归的残差平方和为

$$\frac{\begin{vmatrix} \boldsymbol{y}\boldsymbol{y}' & \boldsymbol{y}\boldsymbol{U}' \\ \boldsymbol{U}\boldsymbol{y}' & \boldsymbol{U}\boldsymbol{U}' \end{vmatrix}}{|\boldsymbol{U}\boldsymbol{U}'|}. \tag{9}$$

证明 由附录中的推论 A.3.1 可知 (9) 为 $\boldsymbol{y}\boldsymbol{y}' - \boldsymbol{y}\boldsymbol{U}'(\boldsymbol{U}\boldsymbol{U}')^{-1}\boldsymbol{U}\boldsymbol{y}'$, 它是类似于 8.2 节 (13) 式所表示的残差平方和. ■

引理 8.4.4 由 (8) 式定义的 V_i 是 $y_{i1}, \cdots, y_{iN}$ 关于 $\boldsymbol{y}_1^{(i-1)}, \cdots, \boldsymbol{y}_N^{(i-1)}$ 和 $\boldsymbol{Z}$ 回归的残差平方和与 $y_{i1}, \cdots, y_{iN}$ 关于 $\boldsymbol{y}_1^{(i-1)}, \cdots, \boldsymbol{y}_N^{(i-1)}$ 和 $\boldsymbol{Z}_2$ 回归的残差平方和之比.

证明 由推论 A.3.1 可知, V_i 的分子可以写为 [利用 8.2 节的 (13) 式]

$$\frac{|\boldsymbol{G}_i|}{|\boldsymbol{G}_{i-1}|}=\frac{|\boldsymbol{Y}_i\boldsymbol{Y}_i'-\boldsymbol{Y}_i\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z}\boldsymbol{Y}_i'|}{|\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}'-\boldsymbol{Y}_{i-1}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1}\boldsymbol{Z}\boldsymbol{Y}_{i-1}'|} \tag{10}$$

$$=\frac{\left|\begin{matrix}\boldsymbol{Y}_i\boldsymbol{Y}_i' & \boldsymbol{Y}_i\boldsymbol{Z}'\\ \boldsymbol{Z}\boldsymbol{Y}_i' & \boldsymbol{Z}\boldsymbol{Z}'\end{matrix}\right|\Big/|\boldsymbol{Z}\boldsymbol{Z}'|}{\left|\begin{matrix}\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}' & \boldsymbol{Y}_{i-1}\boldsymbol{Z}'\\ \boldsymbol{Z}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}\boldsymbol{Z}'\end{matrix}\right|\Big/|\boldsymbol{Z}\boldsymbol{Z}'|}$$

$$=\frac{\left|\begin{matrix}\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}' & \boldsymbol{Y}_i\boldsymbol{y}_i^{*\prime} & \boldsymbol{Y}_{i-1}\boldsymbol{Z}'\\ \boldsymbol{y}_i^*\boldsymbol{Y}_{i-1}' & \boldsymbol{y}_i^*\boldsymbol{y}_i^{*\prime} & \boldsymbol{y}_i^*\boldsymbol{Z}'\\ \boldsymbol{Z}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}\boldsymbol{y}_i^{*\prime} & \boldsymbol{Z}\boldsymbol{Z}'\end{matrix}\right|}{\left|\begin{matrix}\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}' & \boldsymbol{Y}_{i-1}\boldsymbol{Z}'\\ \boldsymbol{Z}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}\boldsymbol{Z}'\end{matrix}\right|}$$

$$=\frac{\left|\begin{matrix}\boldsymbol{y}_i^*\boldsymbol{y}_i^{*\prime} & \boldsymbol{y}_i^*\begin{bmatrix}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}'\end{bmatrix}\\ \begin{bmatrix}\boldsymbol{Y}_{i-1}\\ \boldsymbol{Z}\end{bmatrix}\boldsymbol{y}_i^{*\prime} & \begin{bmatrix}\boldsymbol{Y}_{i-1}\\ \boldsymbol{Z}\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}'\end{bmatrix}\end{matrix}\right|}{\left|\begin{bmatrix}\boldsymbol{Y}_{i-1}\\ \boldsymbol{Z}\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}'\end{bmatrix}\right|}$$

$$=\boldsymbol{y}_i^*\boldsymbol{y}_i^{*\prime}-\boldsymbol{y}_i^*\begin{bmatrix}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}'\end{bmatrix}\begin{bmatrix}\boldsymbol{Y}_{i-1}\boldsymbol{Y}_{i-1}' & \boldsymbol{Y}_{i-1}\boldsymbol{Z}'\\ \boldsymbol{Z}\boldsymbol{Y}_{i-1}' & \boldsymbol{Z}\boldsymbol{Z}'\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_{i-1}\\ \boldsymbol{Z}\end{bmatrix}\boldsymbol{y}_i^{*\prime}.$$

由引理 8.4.3 可知 (10) 式的右端为 $\boldsymbol{y}_i^*$ 关于 $\boldsymbol{Y}_{i-1}$ 和 $\boldsymbol{Z}$ 回归的残差平方和. 分母的计算类似, 只需把 $\boldsymbol{Z}$ 换成 $\boldsymbol{Z}_2$ 即可. ■

比值 V_i 是检验假设 $\boldsymbol{y}_i^*=\boldsymbol{x}_i^*-\boldsymbol{\beta}_{i1}^*\boldsymbol{Z}_1$ 关于 $\boldsymbol{Z}_1$ 的回归等于 0($\boldsymbol{Y}_{i-1}$ 对 $\boldsymbol{Z}_2$ 的回归存在时) 的似然比准则的 $N/2$ 次幂, $\boldsymbol{\beta}_{i1}^*$ 是 $\boldsymbol{\mathrm{B}}_1^*$ 的第 i 行. 当 $i=1$ 时, g_{11} 为 $\boldsymbol{y}_1^*=(y_{11},\cdots,y_{1N})$ 关于 $\boldsymbol{Z}$ 回归的残差平方和, $g_{11}+h_{11}$ 是 $\boldsymbol{y}_1^*=(y_{11},\cdots,y_{1N})$ 关于 $\boldsymbol{Z}_2$ 回归的残差平方和. 比值 $g_{11}/(g_{11}+h_{11})$, 可用来近似检验假设 $\boldsymbol{y}_1^*$ 关于 $\boldsymbol{Z}_1$ 回归为 $\mathbf{0}$, 它的分布为 $\chi_n^2/(\chi_n^2+\chi_m^2)$(引理 8.4.2), 即 β 分布 $\beta(v;\frac{1}{2}n,\frac{1}{2}m)$. (见 5.2 节的例子.) 所以 V_i 具有 β 密度

$$\beta\left[v;\frac{1}{2}(n+1-i),\frac{1}{2}m\right] \tag{11}$$
$$=\begin{cases}\dfrac{\Gamma[\frac{1}{2}(n+m+1-i)]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma(\frac{1}{2}m)}v^{\frac{1}{2}(n+1-i)-1}(1-v)^{\frac{1}{2}m-1}, & 0\leqslant v\leqslant 1,\\ 0, \quad 其他.\end{cases}$$

因为该分布不依赖于 $\boldsymbol{Y}_{i-1}$, 所以比值 V_i 与 $\boldsymbol{Y}_{i-1}$ 独立, 从而与 $V_1,\cdots,V_{i-1}$ 独立. 所以 $V_1,\cdots,V_p$ 是相互独立的.

定理 8.4.1 由 (3) 式定义的 U 的分布为乘积 $\prod_{i=1}^p V_i$ 的分布, 其中 $V_1, \cdots, V_p$ 相互独立且 V_i 的密度为 (11).

U 的分布函数可由 $V_1, \cdots, V_p$ 的联合密度在

$$\prod_{i=1}^{p} V_i \leqslant u, \tag{12}$$

上积分得到.

下面我们将说明对于给定的 $N-q_2$, 指标 p 和 q_1 是可以交换的, 即 $U_{p,q_1,N-q_2-q_1} = U_{p,m,n}$ 的分布与 $U_{q_1,p,N-q_2-p} = U_{m,p,n+m-p}$ 的分布相同. 当 $\boldsymbol{\Sigma} = \boldsymbol{I}$ 以及 $\boldsymbol{\beta}_1 = \mathbf{0}$ 时, 在 8.2 节定义的 $\boldsymbol{G}$ 和 $\boldsymbol{W}_1$ 的联合密度为

$$\frac{|\boldsymbol{G}|^{\frac{1}{2}(n-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{G} - \frac{1}{2}\mathrm{tr}\boldsymbol{W}_1\boldsymbol{W}_1'}}{2^{\frac{1}{2}np} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+1-i)] (2\pi)^{\frac{1}{2}mp}}. \tag{13}$$

设 $\boldsymbol{G} + \boldsymbol{W}_1\boldsymbol{W}_1' = \boldsymbol{J} = \boldsymbol{C}\boldsymbol{C}'$ 和 $\boldsymbol{W}_1 = \boldsymbol{C}\boldsymbol{U}$. 则

$$\begin{aligned} U_{p,m,n} &= \frac{|\boldsymbol{G}|}{|\boldsymbol{G} + \boldsymbol{W}_1\boldsymbol{W}_1'|} = \frac{|\boldsymbol{C}\boldsymbol{C}' - \boldsymbol{C}\boldsymbol{U}\boldsymbol{U}'\boldsymbol{C}'|}{|\boldsymbol{C}\boldsymbol{C}'|} = |\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}'| \\ &= \begin{vmatrix} \boldsymbol{I}_p & \boldsymbol{U} \\ \boldsymbol{U}' & \boldsymbol{I}_m \end{vmatrix} = \begin{vmatrix} \boldsymbol{I}_m & \boldsymbol{U}' \\ \boldsymbol{U} & \boldsymbol{I}_p \end{vmatrix} = |\boldsymbol{I}_m - \boldsymbol{U}'\boldsymbol{U}|; \end{aligned} \tag{14}$$

由附录中的定理 A.3.2 可得第四个和第六个等式, 置换行与列可得第五个等式. 因为 $\boldsymbol{W}_1 = \boldsymbol{C}\boldsymbol{U}$ 的雅可比行列式为 $|\boldsymbol{C}|^m = |\boldsymbol{J}|^{\frac{1}{2}m}$, 当 $\boldsymbol{J}$ 和 $\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}'$ 正定时, $\boldsymbol{J}$ 和 $\boldsymbol{U}$ 的联合密度为

$$\begin{aligned} &\frac{|\boldsymbol{J}|^{\frac{1}{2}(n+m-p-1)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{J}}}{2^{\frac{1}{2}(n+m)p} \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma[\frac{1}{2}(n+m+1-i)]} \\ &\cdot \prod_{i=1}^{p} \left\{ \frac{\Gamma[\frac{1}{2}(n+m+1-i)]}{\Gamma[\frac{1}{2}(n+1-i)]} \right\} \frac{|\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}'|^{\frac{1}{2}(n-p-1)}}{\pi^{\frac{1}{2}mp}}, \end{aligned} \tag{15}$$

否则为 0. 因此 $\boldsymbol{J}$ 和 $\boldsymbol{U}$ 是分布独立的; $\boldsymbol{J}$ 的密度为 (15) 式中的第一项, 即 $w(\boldsymbol{J}|\boldsymbol{I}_p, n+m)$. $\boldsymbol{U}$ 的密度为第二项, 即当 $\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}'$ 正定时为

$$\boldsymbol{K}|\boldsymbol{I}_p - \boldsymbol{U}\boldsymbol{U}'|^{\frac{1}{2}(n-p-1)}, \tag{16}$$

否则为 0. 设 $\boldsymbol{U}_* = \boldsymbol{U}', p^* = m, m^* = p$ 以及 $n^* = n+m-p$. 则当 $\boldsymbol{I}_p - \boldsymbol{U}_*'\boldsymbol{U}_*$ 正定时 $\boldsymbol{U}_*$ 的密度为

$$\boldsymbol{K}|\boldsymbol{I}_p - \boldsymbol{U}_*'\boldsymbol{U}_*|^{\frac{1}{2}(n-p-1)}, \tag{17}$$

否则为 0. 由 (14) 可知 $|\boldsymbol{I}_p - \boldsymbol{U}_*'\boldsymbol{U}_*| = |\boldsymbol{I}_m - \boldsymbol{U}_*\boldsymbol{U}_*'|$, 因此 $\boldsymbol{U}_*$ 的密度为

$$\boldsymbol{K}|\boldsymbol{I}_{p^*} - \boldsymbol{U}_*\boldsymbol{U}_*'|^{\frac{1}{2}(n^*-p^*-1)}. \tag{18}$$

把 (16) 式中的 p 换成 $p^* = m$, m 换成 $m^* = p$, $n-p-1$ 换成 $n^*-p^*-1 = n-p-1$, 就得到 (18) 式. 最后我们指出 (14) 式给出的 $U_{p,m,n}$ 为 $|\boldsymbol{I}_m - \boldsymbol{U}_*\boldsymbol{U}_*'| = U_{m,p,n+m-p}$.

定理 8.4.2 当原假设成立时, $U_{p,q_1,N-q_1-q_2}$ 的分布与 $U_{q_1,p,N-p-q_2}$ 的分布相同 (即 $U_{p,m,n}$ 的分布与 $U_{m,p,n+m-p}$ 的相同).

8.4.2 矩

因为 (11) 是一个密度, 所以积分为 1, 通过改变记法

$$\int_0^1 v^{a-1}(1-v)^{b-1}\mathrm{d}v = B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \tag{19}$$

可得 V_i 的 h 阶矩为

$$\begin{aligned}\mathrm{E}(V_i^h) &= \int_0^1 \frac{\Gamma[\frac{1}{2}(n+m+1-i)]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma(\frac{1}{2}m)} v^{\frac{1}{2}(n+1-i)+h-1}(1-v)^{\frac{1}{2}m-1}\mathrm{d}v \\ &= \frac{\Gamma[\frac{1}{2}(n+1-i)+h]\Gamma[\frac{1}{2}(n+m+1-i)]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma[\frac{1}{2}(n+m+1-i)+h]}.\end{aligned} \tag{20}$$

因为 $V_1,\cdots,V_p$ 相互独立, 所以 $\mathrm{E}(U^h)=\mathrm{E}(\prod_{i=1}^p V_i^h)=\prod_{i=1}^p \mathrm{E}(V_i^h)$.

定理 8.4.3 *U 的 h [若 $h>-\frac{1}{2}(n+1-p)$] 阶矩为*

$$\begin{aligned}\mathrm{E}(U^h) &= \prod_{i=1}^p \frac{\Gamma[\frac{1}{2}(n+1-i)+h]\Gamma[\frac{1}{2}(n+m+1-i)]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma[\frac{1}{2}(n+m+1-i)+h]} \\ &= \prod_{i=1}^p \frac{\Gamma[\frac{1}{2}(N-q_1-q_2+1-i)+h]\Gamma[\frac{1}{2}(N-q_2+1-i)]}{\Gamma[\frac{1}{2}(N-q_1-q_2+1-i)]\Gamma[\frac{1}{2}(N-q_2+1-i)+h]}.\end{aligned} \tag{21}$$

其中第一个展开式中的 p 可以由 m 代替, m 可以由 p 代替, n 可以由 $n+m-p$ 代替.

设 p 为偶数, 即 $p=2r$. 利用倍量公式

$$\Gamma(\alpha+\frac{1}{2})\Gamma(\alpha+1)=\frac{\sqrt{\pi}\Gamma(2\alpha+1)}{2^{2\alpha}}. \tag{22}$$

$U_{2r,m,n}$ 的 h 阶矩为

$$\begin{aligned}\mathrm{E}(U_{2r,m,n}^h) &= \prod_{j=1}^r \left\{ \frac{\Gamma[\frac{1}{2}(m+n+2)-j]}{\Gamma[\frac{1}{2}(m+n+2)-j+h]} \frac{\Gamma[\frac{1}{2}(m+n+1)-j]}{\Gamma[\frac{1}{2}(m+n+1)-j+h]} \right. \\ &\quad \left. \cdot \frac{\Gamma[\frac{1}{2}(n+2)-j+h]\Gamma[\frac{1}{2}(n+1)-j+h]}{\Gamma[\frac{1}{2}(n+2)-j]\Gamma[\frac{1}{2}(n+1)-j]} \right\} \\ &= \prod_{j=1}^r \left\{ \frac{\Gamma(m+n+1-2j)\Gamma(n+1-2j+2h)}{\Gamma(m+n-1-2j+2h)\Gamma(n+1-2j)} \right\}.\end{aligned} \tag{23}$$

由 β 函数的定义知 (23) 式为

$$\begin{aligned}&\prod_{j=1}^r \left\{ \int_0^1 \frac{\Gamma(m+n+1-2j)}{\Gamma(n+1-2j)\Gamma(m)} y^{(n+1-2j)+2h-1}(1-y)^{m-1}\mathrm{d}y \right\} \\ &= \prod_{j=1}^r \mathrm{E}(Y_j^{2h}) = \mathrm{E}\left(\prod_{j=1}^r Y_j^2\right)^h,\end{aligned} \tag{24}$$

其中 Y_j 是独立的并且具有密度 $\beta(y; n+1-2j, m)$.

若 p 是奇数, 即 $p = 2s+1$. 则

$$\mathrm{E}(U_{2s+1,m,n}^h) = \mathrm{E}\left(\prod_{i=1}^{s} Z_i^2 Z_{s+1}\right)^h, \tag{25}$$

其中 $Z_i\,(i=1,\cdots,s)$ 是独立的并且密度为 $\beta(z; n+1-2i, m)$, Z_{s+1} 的密度为 $\beta[z; (n+1-p)/2, m/2]$.

定理 8.4.4　$U_{2r,m,n}$ 与 $\prod_{i=1}^r Y_i^2$ 同分布, 其中 $Y_1, \cdots, Y_r$ 相互独立且 Y_i 的密度为 $\beta(y; n+1-2i, m)$; $U_{2s+1,m,n}$ 与 $\prod_{i=1}^s Z_i^2 Z_{s+1}$ 同分布, 其中 $Z_i\,(i=1,\cdots,s)$ 相互独立并且密度为 $\beta(z; n+1-2i, m)$, Z_{s+1} 与它们也是独立的且密度为 $\beta[z; \frac{1}{2}(n+1-p), \frac{1}{2}m]$.

8.4.3　一些特殊的分布

$p = 1$

由前面的讨论可知 $U_{1,m,n}$ 的密度为

$$\frac{\Gamma[\frac{1}{2}(n+m)]}{\Gamma(\frac{1}{2}n)\Gamma(\frac{1}{2}m)} u^{\frac{1}{2}n-1}(1-u)^{\frac{1}{2}m-1}. \tag{26}$$

$U_{1,m,n}$ 的另一种写法为

$$U_{1,m,n} = \frac{1}{1+\sum_{i=1}^m Y_i^2/g_{11}} = \frac{1}{1+(m/n)F_{m,n}}, \tag{27}$$

其中 g_{11} 为 $\boldsymbol{G} = N\hat{\boldsymbol{\Sigma}}_\Omega$ 中的一个元素, $F_{m,n}$ 为 F 统计量. 因此

$$\frac{1-U_{1,m,n}}{U_{1,m,n}} \cdot \frac{n}{m} = F_{m,n}. \tag{28}$$

定理 8.4.5　$[(1-U_{1,m,n})/U_{1,m,n}]n/m$ 服从自由度为 m 和 n 的 F 分布, $[(1-U_{p,1,n})/U_{p,1,n}](n+1-p)/p$ 服从自由度为 p 和 $n+1-p$ 的 F 分布.

$p = 2$

由定理 8.4.4 可知 $\sqrt{U_{2,m,n}}$ 的密度为

$$\frac{\Gamma(n+m-1)}{\Gamma(n-1)\Gamma(m)} x^{n-2}(1-x)^{m-1}, \tag{29}$$

从而 $U_{2,m,n}$ 的密度为

$$\frac{\Gamma(n+m-1)}{2\Gamma(n-1)\Gamma(m)} u^{\frac{1}{2}(n-3)}(1-\sqrt{u})^{m-1}. \tag{30}$$

由 (29) 可得

$$\frac{1-\sqrt{U_{2,m,n}}}{\sqrt{U_{2,m,n}}} \cdot \frac{n-1}{m} = F_{2m,2(n-1)}. \tag{31}$$

定理 8.4.6 $[(1-\sqrt{U_{2,m,n}})/\sqrt{U_{2,m,n}}]\cdot(n-1)/m$ 服从自由度为 $2m$ 和 $2(n-1)$ 的 F 分布, $[(1-\sqrt{U_{p,2,n}})/\sqrt{U_{p,2,n}}]\cdot(n+1-p)/p$ 服从自由度为 $2p$ 和 $2(n+1-p)$ 的 F 分布.

p 为偶数

Wald and Brookner(1941) 给出了当 p 或 m 为偶数时求解 $U_{p,m,n}$ 的分布的一种方法. 我们将给出 Schatzoff (1966a) 的方法. 为了方便首先考虑 $U_{p,m,n}$, 其中 $m=2r$ 的情况. 把事件 $\prod_{i=1}^{p}V_i\leqslant u$ 写为

$$Y_1+\cdots+Y_p\geqslant-\ln u, \tag{32}$$

其中 $Y_1,\cdots,Y_p$ 是相互独立的, 当 $0\leqslant y<\infty$ 时, $Y_i=-\ln V_i$ 的密度为

$$K_i\mathrm{e}^{-\frac{1}{2}(n+1-i)y}(1-\mathrm{e}^{-y})^{r-1}=K_i\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}\mathrm{e}^{-[\frac{1}{2}(n+1-i)+j]y}, \tag{33}$$

否则为 0, 且

$$K_i=\frac{\Gamma[\frac{1}{2}(n+1-i)+r]}{\Gamma[\frac{1}{2}(n+1-i)]\Gamma(r)}=\frac{1}{(r-1)!}\prod_{j=0}^{r-1}\frac{n+1-i+2j}{2}. \tag{34}$$

则 $Y_1,\cdots,Y_p$ 的联合密度为 $\exp[-\sum_{i=1}^{p}a_iy_i]$ 的线性组合. $W_j=\sum_{i=1}^{j}Y_i$ 的密度可以由 $W_{j-1}=\sum_{i=1}^{j-1}Y_i$ 和 $Y_j, j=2,\cdots,p$ 的密度归纳而得, 它是 $w_{j-1}^k\mathrm{e}^{cw_{j-1}+a_jy_j}$ 的一个线性组合. W_j 的密度由线性组合

$$\begin{aligned}\mathrm{e}^{a_jw_j}\int_0^{w_j}w^k\mathrm{e}^{(c-a_j)w}\mathrm{d}w&=\mathrm{e}^{a_jw_j}\cdot\frac{w_j^{k+1}}{k+1},\quad 若a_j=c,\\&=\mathrm{e}^{cw_j}\sum_{h=0}^{k}(-1)^h\frac{k!}{(k-h)!}\frac{w_j^{k-h}}{(c-a_j)^{h+1}}\\&\quad+(-1)^{k+1}\mathrm{e}^{a_jw_j}\frac{k!}{(c-a_j)^{k+1}},\quad 若a_j\neq c\end{aligned} \tag{35}$$

组成. 计算中涉及部分积分.

定理 8.4.7 *若 p 或 m 是偶数, 则 $U_{p,m,n}$ 的密度可以表示为 $(-\ln u)^ku^l$ 的线性组合, 其中 k 是一个整数, l 是一个半整数.*

由 (35) 我们知道 $-\ln U$ 的累积分布函数是 $w^k\mathrm{e}^{-lw}$ 的线性组合, 因此 U 的累积分布函数是 $(-\ln u)^ku^l$ 的线性组合. k 和 l 的值以及系数依赖于 p,m 和 n. 它们可以通过归纳地应用得到定理 8.4.7 的方法来求得. Pillai and Gupta (1969) 利用定理 8.4.3 的方法得到分布.

另一种方法是使用定理 8.4.4. $U_{2r,m,n}$ 的累积分布函数的补为

$$\begin{aligned}\Pr\{U_{2r,m,n}\geqslant u\}&=\Pr\left\{\prod_{i=1}^{r}Y_i>\sqrt{u}\right\}\\&=\int_{\sqrt{u}}^{1}\int_{\frac{\sqrt{u}}{y_1}}^{1}\cdots\int_{\frac{\sqrt{u}}{\prod_{i=1}^{r-1}y_i}}^{1}\prod_{i=1}^{r}\beta(y_i|n+1-2i,m)\mathrm{d}y_r\cdots\mathrm{d}y_2\mathrm{d}y_1.\end{aligned} \tag{36}$$

在密度中, $(1-y_i)^{m-1}$ 可以通过二项式定理展开. 所以所有的积分都可以表示为变量幂次的积分.

作为一个例子, 我们来讨论 $r=2$ 时的情况. Y_1 和 Y_2 的密度为

$$\begin{aligned} &Cy_1^{n-2}y_2^{n-4}(1-y_1)^{m-1}(1-y_2)^{m-1} \\ =&C\sum_{i,j=0}^{m-1}\frac{[(m-1)!]^2(-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!}y_1^{n-2+i}y_2^{n-4+j}, \end{aligned} \tag{37}$$

其中

$$C=\frac{\Gamma(n+m-1)\Gamma(n+m-3)}{\Gamma(n-1)\Gamma(n-3)\Gamma^2(m)}. \tag{38}$$

$U_{4,m,n}$ 的分布函数的补为

$$\begin{aligned} \Pr\{U_{4,m,n}\geqslant u\}=&C\sum_{i,j=0}^{m-1}\frac{[(m-1)!]^2(-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!} \\ &\cdot\int_{\sqrt{u}}^1\int_{\sqrt{u}y_1}^1 y_1^{n-2+i}y_2^{n-4+j}\mathrm{d}y_2\mathrm{d}y_1 \\ =&C\sum_{i,j=0}^{m-1}\frac{[(m-1)!]^2(-1)^{i+j}}{(m-i-1)!(m-j-1)!i!j!(n-3+j)} \\ &\cdot\int_{\sqrt{u}}^1[y_1^{n-2+i}-u^{\frac{1}{2}(n-3+j)}y_1^{1+i-j}]\mathrm{d}y_1. \end{aligned} \tag{39}$$

最后一个积分可以得到 $\sqrt{u}$ 的幂次, 幂次的乘积以及 $\ln u$ $(1+i-j=-1)$.

特殊值

Wilks (1935) 给出了 U 的几个分布 (当 $m=3$, $p=1,p=2,p=3$; $m=4$, $p=3$; $m=4$, $p=4$ 时). $m=4$, $p=3$ 时 Wilks 的公式中有点错误, 参见该书的第 1 版. Consul (1966) 给出了许多特殊情况下的分布. 参见 Mathai (1971).

8.4.4 似然比过程

令 $u_{p,m,n}(\alpha)$ 是 $U_{p,m,n}$ 在显著性水平为 α 时的上分位点, 即

$$\Pr\{U_{p,m,n}\leqslant u_{p,m,n}(\alpha)|H\ 真\}=\alpha. \tag{40}$$

由 8.5 节可知 $-[n-\frac{1}{2}(p-m+1)]\ln U_{p,m,n}$ 的极限分布是自由度为 pm 的 χ^2 分布. 令 $\chi^2_{pm}(\alpha)$ 代表 χ^2_{pm} 在显著性水平为 α 时的上分位点,

$$C_{p,m,n-p+1}(\alpha)=\frac{-[n-\frac{1}{2}(p-m+1)]\ln u_{p,m,n}(\alpha)}{\chi^2_{pm}(\alpha)}. \tag{41}$$

表 B.1 [Pearson and Hartley (1972)] 给出了 $\alpha=0.1,\alpha=0.05$, $p=1(1)10$, m 为多个偶数, $M=n-p+1=1(1)10(2)20,24,30,40,60,120$ 时的 $C_{p,m,M}(\alpha)$ 的值.

为了检验原假设, 计算出 $U_{p,m,n}$ 的值. 如果下式成立,

$$-\left[n-\frac{1}{2}(p-m+1)\right]\ln U_{p,m,n} > C_{p,m,n-p+1}(\alpha)\chi^2_{pm}(\alpha), \tag{42}$$

则在著性水平 α 下拒绝原假设. 因为 $C_{p,m,n}(\alpha)>1$, 所以 (42) 式左边小于 $\chi^2_{pm}(\alpha)$ 时接受原假设.

之所以将 $C_{p,m,M}(\alpha)$ 列表是因为线性插值比较精确, 其值随着 M 的增大单调递减且光滑地趋向于 1. 通过与偶数个 p 的情况相比较, Schatzoff (1966a) 推荐使用奇数个 p 的插值. 这张表同样显示出 χ^2 近似的精确性. Pillai and Gupta (1969) 对该表进行了拓展.

8.4.5 逐步下降过程

准则 U 在 (7) 中被表示成一些独立 β 变量 $V_1,V_2,\cdots,V_p$ 的乘积. 比值 V_i 为检验原假设 "在 $\boldsymbol{x}_i^*-\beta_{i1}^*\boldsymbol{Z}_1$ 关于 $\boldsymbol{Z}=(\boldsymbol{Z}_1' \quad \boldsymbol{Z}_2')'$ 和 $\boldsymbol{X}_{i-1}$ 的回归中, $\boldsymbol{Z}_1$ 的系数为 $\boldsymbol{0}$" 的最小二乘准则. $\boldsymbol{X}$ 关于 $\boldsymbol{Z}_1$ 回归的原假设为 $\boldsymbol{\beta}_1^*$, 等价于假设 $\boldsymbol{X}-\boldsymbol{\beta}_1^*\boldsymbol{Z}_1$ 关于 $\boldsymbol{Z}_1$ 的回归为 0, 它由假设 $\boldsymbol{x}_i^*-\beta_{i1}^*\boldsymbol{Z}_1\,(i=1,\cdots,p)$ 关于 $\boldsymbol{Z}_1$ 的回归为 0 组成. 因此原假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 可以通过 $V_1,V_2,\cdots,V_p$ 来检验.

因为在假设 $\beta_{i1}=\beta_{i1}^*$ 下, V_i 的密度为 (11), 所以

$$\frac{1-V_i}{V_i}\frac{n-i+1}{m} \tag{43}$$

的分布为自由度为 m 和 $n-i+1$ 的 F 分布. 逐步下降检验过程, 就是将 (43) 式在 $i=1$ 时的值与分位数 $F_{m,n}(\varepsilon_1)$ 进行比较. 若 (43) 在 $i=1$ 时的值比较大, 就拒绝 $\boldsymbol{x}_i^*-\beta_{i1}^*\boldsymbol{Z}_1$ 关于 $\boldsymbol{Z}_1$ 的回归为 0 的原假设, 因此拒绝原假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$. 若第 1 分量的原假设被接受, 比较 $i=2$ 的 (43) 和 $F_{m,n-1}(\varepsilon_2)$. 依次检验各分量的原假设, 若某一个被拒绝, 则检验停止, 假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 被拒绝. 如果所有的关于各分量的原假设都被接受, 则复合假设被接受. 若假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 为真, 则接受概率为 $\prod_{i=1}^p(1-\varepsilon_i)$. 因此逐步下降检验的显著性水平为 $1-\prod_{i=1}^p(1-\varepsilon_i)$.

在逐步下降过程中, 研究者经常会面对变量顺序[①] (即$\boldsymbol{X}$的分量的排序) 的选择问题和分量的显著性水平的选择问题. 把变量按着重要性大小的顺序进行排序似乎比较合理. 显著性水平的选择将影响到功效. 若 ε_i 比较小, 则它将会使得与第 i 个原假设有较大的偏离, 从而导致拒绝. 若不考虑其他的原因, 各分量可以取相等的显著性水平. 当然这种方法关于独立向量变量的线性变换不是不变的. 然而, 在使用逐步下降过程之前, 可以先使用线性变换决定这 p 个变量.

---

① 在某些情况, 变量的顺序是给定的; 例如, x_1 是第一时间点的观测, x_2 为第二时间点的观测, 等等.

可以把因子分组. 例如, 把 $x_1, \cdots, x_k$ 放在一个集合内, $x_{k+1}, \cdots, x_p$ 放在另一个集合内. 则 $U_{k,m,n} = \prod_{i=1}^{k} V_i$ 可以用来检验原假设 "$\boldsymbol{\beta}_1$ 的前 k 行就是 $\boldsymbol{\beta}_1^*$ 的前 k 行". 接着 $\prod_{i=k+1}^{p} V_i$ 可以用来检验原假设 "$\boldsymbol{\beta}_1$ 的后 $p-k$ 行就是 $\boldsymbol{\beta}_1^*$ 的后 $p-k$ 行"; 在原假设下, 后面这个准则的分布为 $U_{p-k,m,n-k}$.

也可以采用似然比过程检验原假设 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$. 如果假设被拒绝, 可以通过研究因子 $V_1, V_2, \cdots, V_p$, 来判断 $\boldsymbol{\beta}_1$ 的哪些行与 $\boldsymbol{\beta}_1^*$ 的不同.

也可以利用这些因子去获得 $\boldsymbol{\beta}_{11}, \cdots, \boldsymbol{\beta}_{p1}$ 的置信区域. 设 $v_i(\varepsilon_i)$ 由下式定义,

$$\frac{1 - v_i(\varepsilon_i)}{v_i(\varepsilon_i)} \frac{n-i+1}{m} = F_{m,n-i+1}(\varepsilon_i). \tag{44}$$

置信度为 $1 - \varepsilon_i$ 的 $\boldsymbol{\beta}_{i1}$ 的置信区域为

$$\frac{\begin{vmatrix} \boldsymbol{x}_i^* \boldsymbol{x}_i^{*\prime} & \boldsymbol{x}_i^* \boldsymbol{X}_{i-1}' & \boldsymbol{x}_i^* \boldsymbol{Z}' \\ \boldsymbol{X}_{i-1} \boldsymbol{x}_i^{*\prime} & \boldsymbol{X}_{i-1} \boldsymbol{X}_{i-1}' & \boldsymbol{X}_{i-1} \boldsymbol{Z}' \\ \boldsymbol{Z} \boldsymbol{x}_i^{*\prime} & \boldsymbol{Z} \boldsymbol{X}_{i-1}' & \boldsymbol{Z} \boldsymbol{Z}' \end{vmatrix}}{\begin{vmatrix} (\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1)(\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1)' & (\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1) \boldsymbol{X}_{i-1}' & (\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1) \boldsymbol{Z}_2' \\ \boldsymbol{X}_{i-1} (\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1)' & \boldsymbol{X}_{i-1} \boldsymbol{X}_{i-1}' & \boldsymbol{X}_{i-1} \boldsymbol{Z}_2' \\ \boldsymbol{Z}_2 (\boldsymbol{x}_i^* - \bar{\boldsymbol{\beta}}_{i1} \boldsymbol{Z}_1)' & \boldsymbol{Z}_2 \boldsymbol{X}_{i-1}' & \boldsymbol{Z}_2 \boldsymbol{Z}_2' \end{vmatrix}} \cdot \frac{\begin{vmatrix} \boldsymbol{X}_{i-1} \boldsymbol{X}_{i-1}' & \boldsymbol{Z}_{i-1} \boldsymbol{Z}_2^* \\ \boldsymbol{Z}_2 \boldsymbol{X}_{i-1}' & \boldsymbol{Z}_2 \boldsymbol{X}_2' \end{vmatrix}}{\begin{vmatrix} \boldsymbol{X}_{i-1} \boldsymbol{X}_{i-1}' & \boldsymbol{X}_{i-1} \boldsymbol{Z}' \\ \boldsymbol{Z} \boldsymbol{X}_{i-1}' & \boldsymbol{Z} \boldsymbol{Z}' \end{vmatrix}} \geqslant v_i(\varepsilon_i). \tag{45}$$

8.5 似然比准则的分布的渐近展开

8.5.1 渐近展开的一般理论

本节我们将研究关于本章准则的大样本分布理论. 首先我们讨论某类随机变量分布的一般渐近展开, 其中该随机变量的矩为 Γ 函数 [Box (1949)] 的函数. 接着我们把这种渐近展开应用到线性假设的似然比准则.

考虑 h 阶矩①为

$$\mathrm{E}(\boldsymbol{W}^h) = K \left(\frac{\prod_{j=1}^{b} y_j^{y_j}}{\prod_{k=1}^{a} x_k^{x_k}} \right)^h \frac{\prod_{k=1}^{a} \Gamma[x_k(1+h) + \xi_k]}{\prod_{j=1}^{b} \Gamma[y_j(1+h) + \eta_j]}, \quad h = 0, 1, \cdots \tag{1}$$

的随机变量 $\boldsymbol{W}$ $(0 \leqslant \boldsymbol{W} \leqslant 1)$, 其中 K 为满足 $\mathrm{E}(\boldsymbol{W}^0) = 1$ 和

$$\sum_{k=1}^{a} x_k = \sum_{j=1}^{b} y_j \tag{2}$$

① 在所有我们应用到该结果的情形下, 参数 x_k, ξ_k, y_j, η_j 都服从一个具有这样矩的分布.

的常量.

可以看出, $\lambda=U_{p,q_1,n}^{\frac{1}{2}N}$ 的 h 阶矩是这种形式, 其中 $x_k=\frac{1}{2}N=y_j,\xi_k=\frac{1}{2}(-q+1-k),\eta_j=\frac{1}{2}(-q_2+1-j),a=b=p$. 由于本书后面应用的需要, 我们考虑更一般的情形.

假设

$$\boldsymbol{M}=-2\ln\boldsymbol{W},\tag{3}$$

$\rho\boldsymbol{M}\ (0\leqslant\rho<1)$ 的特征函数为

$$\begin{aligned}\phi(t)&=\mathrm{E}(\mathrm{e}^{\mathrm{i}t\rho\boldsymbol{M}})=\mathrm{E}(\boldsymbol{W}^{-2\mathrm{i}t\rho})\\&=K\left(\frac{\prod_{j=1}^{b}y_j^{y_j}}{\prod_{k=1}^{a}x_k^{x_k}}\right)^{-2\mathrm{i}t\rho}\frac{\prod_{k=1}^{a}\Gamma[x_k(1-2\mathrm{i}t\rho)+\xi_k]}{\prod_{j=1}^{b}\Gamma[y_j(1-2\mathrm{i}t\rho)+\eta_j]}.\end{aligned}\tag{4}$$

这里 ρ 是任意的, 后面它将依赖于 N 的大小. 若 $a=b,x_k=y_k,\xi_k\leqslant\eta_k$, 则 (1) 为具有 β 分布的变量的幂之乘积的 h 阶矩. 对任意的 h, 只要 Γ 函数存在, (1) 式就成立. 此时对任意实值 t, (4) 都是有意义的. 在这里我们假设 (4) 对任意的实值 t 都是成立的, 每当使用这个结果时都要验证这个假设.

设

$$\Phi(t)=\ln\phi(t)=g(t)-g(0),\tag{5}$$

其中

$$\begin{aligned}g(t)=&\,2\mathrm{i}t\rho\left(\sum_{k=1}^{a}x_k\ln x_k-\sum_{j=1}^{b}y_j\ln y_j\right)\\&+\prod_{k=1}^{a}\ln\Gamma[\rho x_k(1-2\mathrm{i}t)+\beta_k+\xi_k]\\&-\prod_{j=1}^{b}\ln\Gamma[\rho y_j(1-2\mathrm{i}t)+\varepsilon_j+\eta_j],\end{aligned}$$

$\beta_k=(1-\rho)x_k$ 和 $\varepsilon_j=(1-\rho)y_j$. 由 $g(t)-g(0)$ 得 $\Phi(0)=0$, 这与 K 满足 $\phi(0)=1$ 的事实相一致. 利用 Γ 函数 [Barnes (1899), p. 64] 的展开公式, 它是以界 h 关于 x 渐近的,

$$\begin{aligned}\ln\Gamma(x+h)=&\ln\sqrt{2\pi}+\left(x+h-\frac{1}{2}\right)\ln x-x\\&-\sum_{r=1}^{m}(-1)^r\frac{B_{r+1}(h)}{r(r+1)x^r}+R_{m+1}(x),\end{aligned}\tag{6}$$

其中 $R_{m+1}(x)=O(x^{-(m+1)})$①, $B_r(h)$ 为 r 次的伯努利多项式, 顺序单位由

① $R_{m+1}(x)=O(x^{-(m+1)})$ 表示当 $|x|\to\infty$ 时, $|x^{m+1}R_{m+1}|$ 是有界的.

$$\frac{\tau e^{\tau h}}{e^{\tau}-1}=\sum_{r=0}^{\infty}\frac{\tau^r}{r!}B_r(h) \tag{7}$$

定义①. 前三个多项式为 $[B_0(h)=1]$

$$\begin{aligned}B_1(h)&=h-\frac{1}{2},\\ B_2(h)&=h^2-h+\frac{1}{6},\\ B_3(h)&=h^3-\frac{3}{2}h^2+\frac{1}{2}h.\end{aligned} \tag{8}$$

依次取 $x=\rho x_k(1-2\mathrm{i}t),\rho y_j(1-2\mathrm{i}t)$ 和 $h=\beta_k+\xi_k,\varepsilon_j+\eta_j$, 可得

$$\begin{aligned}\Phi(t)=&Q-g(0)-\frac{1}{2}f\ln(1-2\mathrm{i}t)\\ &+\sum_{r=1}^{m}\omega_r(1-2\mathrm{i}t)^{-r}+\sum_{k=1}^{a}O(x_k^{-(m+1)})+\sum_{j=1}^{b}O(y_j^{-(m+1)}),\end{aligned} \tag{9}$$

其中

$$f=-2\left\{\sum_k\xi_k-\sum_j\eta_j-\frac{1}{2}(a-b)\right\}, \tag{10}$$

$$\omega_r=\frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_k\frac{B_{r+1}(\beta_k+\xi_k)}{(\rho x_k)^r}-\sum_j\frac{B_{r+1}(\varepsilon_j+\eta_j)}{(\rho y_j)^r}\right\}, \tag{11}$$

$$\begin{aligned}Q=&\frac{1}{2}(a-b)\ln 2\pi-\frac{1}{2}f\ln\rho\\ &+\sum_k\left(x_k+\xi_k-\frac{1}{2}\right)\ln x_k-\sum_j\left(y_j+\eta_j-\frac{1}{2}\right)\ln y_j.\end{aligned} \tag{12}$$

$\phi(t)$ 的另一种表示为 (在这里我们用不到)

$$\phi(t)=e^{\Phi(t)}=e^{Q-g(0)}(1-2\mathrm{i}t)^{-\frac{1}{2}f}\sum_{v=0}^{m}a_v(1-2\mathrm{i}t)^{-v}+R_{m+1}^*, \tag{13}$$

其中 $\sum_{v=0}^{m}a_vz^{-v}$ 为 $\exp(-\sum_{r=0}^{m}\omega_rz^{-r})$ 的级数展开的前 $m+1$ 项的和, R_{m+1}^* 是余项. 或者

$$\Phi(t)=-\frac{1}{2}f\ln(1-2\mathrm{i}t)+\sum_{r=1}^{m}\omega_r[(1-2\mathrm{i}t)^{-r}-1]+R'_{m+1}, \tag{14}$$

其中

$$R'_{m+1}=\sum_k O(x_k^{-(m+1)})+\sum_j O(y_j^{-(m+1)}). \tag{15}$$

---

① 这个定义与 Whittaker and Watson[(1943), p.126] 的有些不同, 后者展开 $\tau(e^{\tau h}-1)/(e^{\tau}-1)$. 若 $B_r^*(h)$ 为第二类多项式, 则 $B_1(h)=B_1^*(h)-\frac{1}{2}, B_{2r}(h)=B_{2r}^*(h)+(-1)^{r+1}B_r$, 其中 B_r 为第 r 个伯努利数, 并且 $B_{2r+1}(h)=B_{2r+1}^*(h)$.

在 (14) 式中, 以展开 $g(t)$ 的方式展开 $g(0)$, 并合并同类项.

那么

$$\begin{aligned}\phi(t) &= \mathrm{e}^{\Phi(t)} \qquad (16)\\ &= (1-2\mathrm{i}t)^{-\frac{1}{2}f}\exp\left[\sum_{r=1}^{m}\omega_r(1-2\mathrm{i}t)^{-r}-\sum_{r=1}^{m}\omega_r+R'_{m+1}\right]\\ &= (1-2\mathrm{i}t)^{-\frac{1}{2}f}\left\{\prod_{r=1}^{m}\left[1+\omega_r(1-2\mathrm{i}t)^{-r}+\frac{1}{2!}\omega_r^2(1-2\mathrm{i}t)^{-2r}\cdots\right]\right.\\ &\quad\left.\times\prod_{r=1}^{m}\left(1-\omega_r+\frac{1}{2!}\omega_r^2-\cdots\right)+R''_{m+1}\right\}\\ &= (1-2\mathrm{i}t)^{-\frac{1}{2}f}[1+T_1(t)+T_2(t)+\cdots+T_m(t)+R'''_{m+1}],\end{aligned}$$

其中 $T_r(t)$ 表示展开式中关于 $\omega_1^{s_1}\cdots\omega_r^{s_r},\sum is_i=r$ 的项; 例如

$$T_1(t)=\omega_1[(1-2\mathrm{i}t)^{-1}-1], \qquad (17)$$

$$T_2(t)=\omega_2[(1-2\mathrm{i}t)^{-2}-1]+\frac{1}{2}\omega_1^2[(1-2\mathrm{i}t)^{-2}-2(1-2\mathrm{i}t)^{-1}+1]. \qquad (18)$$

在多数应用中, 我们取 $x_k=c_k\theta$ 和 $y_j=d_j\theta$, 其中 c_k,d_j 是常量, θ 是变量 (即随着样本量增长). 在这种情况下, 如果 ρ 使得 $(1-\rho)x_k$ 和 $(1-\rho)y_j$ 的极限存在, 则 R'''_{m+1} 为 $O(\theta^{-(m+1)})$. 合并 (16) 式中所有关于 $\omega_1^{s_1}\cdots\omega_r^{s_r},\sum is_i=r$ 的项, 因为这些项为 $O(\theta^{-r})$.

我们可以看到 $T_r(t)$ 是 $(1-2\mathrm{i}t)^{-1}$ 中的次数为 r 的多项式且对于整数 v, $(1-2\mathrm{i}t)^{-\frac{1}{2}f}T_r(t)$ 的每一项都是 $(1-2\mathrm{i}t)^{-\frac{1}{2}v}$ 的常数倍. 我们知道 $(1-2\mathrm{i}t)^{-\frac{1}{2}v}$ 是自由度为 v 的 χ^2 密度的特征函数, 即

$$\begin{aligned}g_v(z) &= \frac{1}{2^{\frac{1}{2}v}\Gamma(\frac{1}{2}v)}z^{\frac{1}{2}v-1}\mathrm{e}^{-\frac{1}{2}z} \qquad (19)\\ &= \int_{-\infty}^{\infty}\frac{1}{2\pi}(1-2\mathrm{i}t)^{-\frac{1}{2}v}\mathrm{e}^{-\mathrm{i}tz}\mathrm{d}t.\end{aligned}$$

设

$$\begin{aligned}S_r(z) &= \int_{-\infty}^{\infty}\frac{1}{2\pi}(1-2\mathrm{i}t)^{-\frac{1}{2}f}T_r(t)\mathrm{e}^{-\mathrm{i}tz}\mathrm{d}t,\\ R^{(4)}_{m+1} &= \int_{-\infty}^{\infty}\frac{1}{2\pi}(1-2\mathrm{i}t)^{-\frac{1}{2}f}R'''_{m+1}\mathrm{e}^{-\mathrm{i}tz}\mathrm{d}t.\end{aligned} \qquad (20)$$

则 ρM 的密度为

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{2\pi}\phi(t)\mathrm{e}^{-\mathrm{i}tz}\mathrm{d}t &= \sum_{r=0}^{m}S_r(z)+R^{(4)}_{m+1} \qquad (21)\\ &= g_f(z)+\omega_1[g_{f+2}(z)-g_f(z)]\\ &\quad+\Big\{\omega_2[g_{f+4}(z)-g_f(z)]\end{aligned}$$

$$+\frac{\omega_1^2}{2}[g_{f+4}(z)-2g_{f+2}(z)+g_f(z)]\Big\}$$
$$+\cdots+S_m(z)+R_{m+1}^{(4)}.$$

设

$$\begin{aligned}U_r(z_0)&=\int_0^{z_0}S_r(z)\mathrm{d}z,\\R_{m+1}^{(5)}&=\int_0^{z_0}R_{m+1}^{(4)}\mathrm{d}z.\end{aligned}\tag{22}$$

$\boldsymbol{M}$ 的分布函数是以 $\rho\boldsymbol{M}$ 的分布函数的形式写出的, 即

$$\begin{aligned}&\Pr\{\boldsymbol{M}\leqslant M_0\}\\&=\Pr\{\rho\boldsymbol{M}\leqslant\rho M_0\}\\&=\sum_{r=0}^{m}U_r(\rho M_0)+R_{m+1}^{(5)}\\&=\Pr\{\chi_f^2\leqslant\rho M_0\}+\omega_0\Big(\Pr\{\chi_{f+2}^2\leqslant\rho M_0\}-\Pr\{\chi_f^2\leqslant\rho M_0\}\Big)\\&\quad+\Big[\omega_2\Big(\Pr\{\chi_{f+4}^2\leqslant\rho M_0\}-\Pr\{\chi_f^2\leqslant\rho M_0\}\Big)+\frac{\omega_1^2}{2}\Big(\Pr\{\chi_{f+4}^2\leqslant\rho M_0\}\\&\quad-2\Pr\{\chi_{f+2}^2\leqslant\rho M_0\}+\Pr\{\chi_f^2\leqslant\rho M_0\}\Big)\Big]\\&\quad+\cdots+U_m(\rho M_0)+R_{m+1}^{(5)}.\end{aligned}\tag{23}$$

剩余项 $R_{m+1}^{(5)}$ 为 $O(\theta^{-(m+1)})$, 最后的陈述可以通过剩余项证明. (事实上, 为了严格证明该结论, 我们只需证明在一致的意义下每一个剩余项具有合适的阶.)

在许多情况下我们希望选择满足 $\omega_1=0$ 的 ρ. 这时, 只需利用 (23) 的第一项就可以得到阶为 θ^{-2} 的误差.

关于展开的详细讨论可以参阅 Box(1949).

定理 8.5.1　*设 (2) 式成立, 对于纯虚数 h, $\mathrm{E}(\boldsymbol{W}^h)$ 由 (1) 式给出. 则 $-2\rho\ln\boldsymbol{W}$ 的分布函数为 (23). 若 $x_k\geqslant c_k\theta, y_j\geqslant d_j\theta(c_k>0,d_j>0)$, 且 $(1-\rho)x_k$ 和 $(1-\rho)y_j$ 有极限, 则误差 $R_{m+1}^{(5)}$ 为 $O(\theta^{-(m+1)})$, 其中 ρ 可能依赖于 θ.*

Box 还考虑了通过 F 分布来近似 $-2\rho\ln\boldsymbol{W}$ 的分布. 他发现该近似的误差可以达到 θ^{-3} 阶.

8.5.2　似然比准则的渐近分布

现在把定理 8.5.1 应用到 8.3 节得到的似然比准则 $-2\ln\lambda$ 的分布. 令 $\boldsymbol{W}=\lambda$. λ 的 h 阶矩为

$$\mathrm{E}(\lambda^h)=K\frac{\prod_{k=1}^{p}\Gamma[\frac{1}{2}(N-q+1-k+Nh)]}{\prod_{j=1}^{p}\Gamma[\frac{1}{2}(N-q_2+1-j+Nh)]},\tag{24}$$

该式对所有 Γ 函数存在的 h 都成立, 包括纯虚数的 h. 设 $a=b=p$,

$$\begin{aligned} x_k &= \frac{1}{2}N, &\quad \xi_k &= \frac{1}{2}(-q+1-k), &\quad \beta_k &= \frac{1}{2}(1-\rho)N, \\ y_j &= \frac{1}{2}N, &\quad \eta_j &= \frac{1}{2}(-q_2+1-j), &\quad \varepsilon_j &= \frac{1}{2}(1-\rho)N. \end{aligned} \tag{25}$$

易见

$$\begin{aligned} 2\omega_1 = &\prod_{k=1}^{p}\left\{\frac{\{\frac{1}{2}[(1-\rho)N-q+1-k]\}^2-\frac{1}{2}[(1-\rho)N-q+1-k]}{\frac{1}{2}\rho N}\right. \\ &\left.-\frac{\{\frac{1}{2}[(1-\rho)N-q_2+1-k]\}^2-\frac{1}{2}[(1-\rho)N-q_2+1-k]}{\frac{1}{2}\rho N}\right\} \\ =&\frac{2}{\rho N}\prod_{k=1}^{p}\left[\frac{-2[(1-\rho)N-q_2+1-k]q_1+q_1^2}{4}+\frac{q_1}{2}\right] \\ =&\frac{pq_1}{2\rho N}[-2(1-\rho)N+2q_2-2+(p+1)+q_1+2]. \end{aligned} \tag{26}$$

为了使它为零, 需要

$$\rho=\frac{N-q_2-\frac{1}{2}(p+q_1+1)}{N}. \tag{27}$$

这时

$$\begin{aligned} &\Pr\left\{-2\frac{k}{N}\ln\lambda\leqslant z\right\} \\ =&\Pr\{-k\ln U_{p,q_1,N-q}\leqslant z\} \\ =&\Pr\{\chi^2_{pq_1}\leqslant z\} \\ &+\frac{\gamma_2}{k^2}\Big(\Pr\{\chi^2_{pq_1+4}\leqslant z\}-\Pr\{\chi^2_{pq_1}\leqslant z\}\Big) \\ &+\frac{1}{k^4}\Big[\gamma_4\Big(\Pr\{\chi^2_{pq_1+8}\leqslant z\}-\Pr\{\chi^2_{pq_1}\leqslant z\}\Big) \\ &-\gamma_2^2\Big(\Pr\{\chi^2_{pq_1+4}\leqslant z\}-\Pr\{\chi^2_{pq_1}\leqslant z\}\Big)\Big]+R_5^{(5)}, \end{aligned} \tag{28}$$

其中

$$k=\rho N=N-q_2-\frac{1}{2}(p+q_1+1)=n-\frac{1}{2}(p-q_1+1), \tag{29}$$

$$\gamma_2=\frac{pq_1(p^2+q_1^2-5)}{48}, \tag{30}$$

$$\gamma_4=\frac{\gamma_2^2}{2}+\frac{pq_1}{1920}[3p^4+3q_1^4+10p^2q_1^2-50(p^2+q_1^2)+159]. \tag{31}$$

因为 $\lambda=U_{p,q_1,n}^{\frac{1}{2}N}$, 所以 (28) 给出了 $\Pr\{-k\ln U_{p,q_1,n}\leqslant z\}$, 其中 $n=N-q$.

定理 8.5.2 $-k\ln U_{p,q_1,n}$ **的分布由 (28) 式给出, 其中** $k=n-\frac{1}{2}(p-q_1+1)$, γ_2 **和** γ_4 **分别由 (30) 和 (31) 给出. 余项为** $O(N^{-6})$.

系数 $k = n - \frac{1}{2}(p - q_1 + 1)$ 被称作 Bartlett 修正. 使用 (28) 式的第一项, 则误差的阶为 N^{-2}; 使用第二项, 则误差的阶为 N^{-4}; 使用第三项[①], 则误差的阶为 N^{-6}. 第二项恒为负, 且当 $z = \sqrt{(pq_1+2)(pq_1)}$ $(= pq_1 + 1$, 近似地) 时达到极大. 对于 $p \geqslant 3, q_1 \geqslant 3, \gamma_2/k^2 \leqslant [(p^2+q_1^2)/k]^2/96$, 第二项在 $-0.005[(p^2+q_1^2)/k]^2$ 和 0 之间; 并且 $\gamma_4 \leqslant \gamma_2^2$, 第三项小于 $(\gamma_2/k^2)^2$. 我们可以遵循一个粗糙的规则: 当 $p^2 + q_1^2 \leqslant k/3$ 时, 使用第一项可精确到小数点后第三位.

考虑 $p = 3, q_1 = 6, N - q_2 = 24, z = 26.0$ (χ^2_{18} 的 10% 分位数). 此时 $\gamma_2/k^2 = 0.048$, 且第二项为 -0.007; $\gamma_4/k^4 = 0.0015$, 第三项为 -0.0001. 因此若保留到小数点后三位, $-19 \ln U_{3,6,18} \leqslant 26.0$ 的概率为 0.893.

因为

$$-\left[n - \frac{1}{2}(p - m + 1)\right] \ln u_{p,m,n}(\alpha) = C_{p,m,n-p+1}(\alpha)\chi^2_{pm}(\alpha), \tag{32}$$

用 $\chi^2_{pm}(\alpha)$ 近似上式左边项的比例误差为 $C_{p,m,n-p+1} - 1$. 随着 p 和 m 的增长, 比例误差增长缓慢.

8.5.3　一种正态近似

当 p 和 m 同时趋向于无穷时, 或者两者中有一个趋向于无穷时, Mudholkar and Trivedi (1980), (1981) 得到了分布 $-\ln U_{p,m,n}$ 的一个正态近似. 它与 χ^2 分布的 Wilson-Hilferty 正态近似相关.

首先我们给出近似的背景. 设 $\{\boldsymbol{Y}_k\}$ 为一个非负随机变量序列, 且满足当 $k \to \infty$ 时, $(\boldsymbol{Y}_k - \mu_k)/\sigma_k \xrightarrow{L} N(0,1)$, 其中 $\mathrm{E}(\boldsymbol{Y}_k) = \mu_k$ 和 $\mathrm{Var}(\boldsymbol{Y}_k) = \sigma_k^2$. 假设 $\mu_k \to \infty$, 当 $k \to \infty$ 时 σ_k^2/μ_k 是有界的. 设 $\boldsymbol{Z}_k = (\boldsymbol{Y}_k/\mu_k)^h$. 则由定理 4.2.3 可知

$$\frac{\boldsymbol{Z}_k - 1}{h(\sigma_k/\mu_k)} = \frac{\mu_k(\boldsymbol{Z}_k - 1)}{h\sigma_k} \xrightarrow{L} N(0,1). \tag{33}$$

选择 h 使得 $\boldsymbol{Z}_k$ 的分布接近对称就可以加速其正态化的进程, 而分布的对称性可由三阶累积量来度量. 事实证明用正态分布来近似比较精确, 而且容易将这种思想推广到极限理论, 尽管这种推广有时不一定是必要的.

通过泰勒展开, 我们把 $\boldsymbol{Y}_k/\mu_k$ 的 h 阶矩表示为

$$\begin{aligned}\mathrm{E}(\boldsymbol{Z}_k) &= \mathrm{E}\left(\frac{\boldsymbol{Y}_k}{\mu_k}\right)^h \\ &= 1 + \frac{h(h-1)}{2}\frac{\sigma_k^2}{\mu_k} \\ &\quad + \frac{h(h-1)(h-2)}{24}\frac{4\phi_k - 3(h-3)(\sigma_k^2/\mu_k)^2}{\mu_k^2} + O(\mu_k^{-3}),\end{aligned} \tag{34}$$

① Box 证明了阶为 N^{-5} 的项为 0, 并给出了系数在阶为 N^{-6} 的项中的应用.

其中 $\phi_k = \mathrm{E}(\boldsymbol{Y}_k - \mu_k)^3/\mu_k$, 并设其有界. 把 (34) 中的 h 换成 rh 就可以得到 $\boldsymbol{Z}_k$ 的 r 阶矩. $\boldsymbol{Z}_k$ 的中心矩为

$$\mathrm{E}(\boldsymbol{Z}_k - 1)^2 = h^2 \frac{\sigma_k^2/\mu_k}{\mu_k} + \frac{h^2(h-1)}{2} \frac{2\phi_k + (3h-5)(\sigma_k^2/\mu_k)^2}{\mu_k^2} + O(\mu_k^{-3}), \tag{35}$$

$$\mathrm{E}(\boldsymbol{Z}_k - 1)^3 = h^3 \frac{\phi_k + 3(h-1)(\sigma_k^2/\mu_k)^2}{\mu_k^2} + O(\mu_k^{-3}). \tag{36}$$

为了使三阶矩近似为 0, 我们取 h 为

$$h_0 = 1 - \frac{\mathrm{E}(\boldsymbol{Y}_k - \mu_k)^3 \mu_k}{3\sigma_k^4}. \tag{37}$$

则当 $h = h_0$ 时, 可以把 $\boldsymbol{Z}_k = (\boldsymbol{Y}_k/\mu_k)^{h_0}$ 看作均值为 (34), 方差为 (35) 的正态分布.

现在考虑 $-\ln U_{p,m,n} = -\sum_{i=1}^{p} \ln V_i$, 其中 $V_1, \cdots, V_p$ 相互独立且 V_i 的密度为 $\beta(x; (n+1-i)/2, m/2), i = 1, \cdots, p$. 当 $n \to \infty$ 以及 $m \to \infty$ 时, $-\ln V_i$ 趋向于正态分布. 若 V 的密度为 $\beta(x; a/2, b/2)$, 则 $-\ln V$ 的矩母函数为

$$\mathrm{E}(\mathrm{e}^{-t\ln V}) = \frac{\Gamma[(a+b)/2]\Gamma(a/2-t)}{\Gamma(a/2)\Gamma[(a+b)/2-t]}. \tag{38}$$

它的对数为累积量母函数. 对最后一项求导可得 V 的 r 阶累积量

$$C_r = (-1)^r \left[\psi^{(r-1)}\left(\frac{a}{2}\right) - \psi^{(r-1)}\left(\frac{a+b}{2}\right) \right], \quad r = 1, 2, \cdots, \tag{39}$$

其中 $\psi(w) = \mathrm{d}\ln\Gamma(w)/\mathrm{d}w$. [例如, Abramovitz and Stegun (1972), p.258.] 由 $\Gamma(w+1) = w\Gamma(w)$ 可得递推关系 $\psi(w+1) = \psi(w) + 1/w$. 从而可得 $s = 0$ 和 l 为整数时的结果

$$\psi^{(s)}(w+l) - \psi^{(s)}(w) = (-1)^{s+1} s! \sum_{j=0}^{l-1} \frac{1}{(w+j)^{s+1}}. \tag{40}$$

对于 $s = 1, 2, \cdots$, (40) 式可以通过微分证得. [Mudholkar and Trivedi (1981) 第 223 页的第一行中的 $\psi'(\boldsymbol{Z})$ 的展开是不正确的.] 因此当 $b = 2l$ 时

$$C_r = (r-1)! \sum_{j=0}^{l-1} \frac{1}{(a/2+j)^r}. \tag{41}$$

由这些结果可得 $-\ln U_{p,2l,n}$ 的 r 阶累积量为

$$\kappa_r(-\ln U_{p,2l,n}) = 2^r (r-1)! \sum_{i=1}^{p} \sum_{j=0}^{l-1} \frac{1}{(n-i+1-2j)^r}. \tag{42}$$

当 $l \to \infty$ 时, $r = 1$ 时的级数是发散的, $r = 2, 3$ 时是收敛的, 因此 $\kappa_r/\kappa_1 \to 0, r = 2, 3$. 当 $p \to \infty$(如果 n/p 接近于一个正常数) 时可得类似的结果.

给定 n, p 和 l, 前三阶累积量可以通过 (42) 计算. h_0 由 (37) 决定, 则当 $h = h_0$ 时, $(-\ln U_{p,2l,n})^{h_0}$ 可以看作均值为 (34), 方差为 (35) 的正态分布.

Mudholkar and Trivedi (1980) 在显著性水平 0.01, 0.05 下, 计算了 n 从 4 到 66, $p = 3, 7$ 和 $q = 2, 6, 10$ 时的逼近误差. 极大误差小于 0.0007, 在大多数情况下误差是相当小的. χ^2 近似的误差是比较大的, 特别是对于较小的 n.

当 m 为奇数时, r 阶累积量可以近似为

$$2^r(r-1)!\sum_{i=1}^{p}\left[\sum_{j=0}^{\frac{1}{2}(m-3)}\frac{1}{(n-i+1-2j)^r}+\frac{1}{2}\frac{1}{(n-i+m)^r}\right]. \tag{43}$$

Davis (1933), (1935) 给出了 $\psi(w)$ 和它导数的列表.

8.5.4　一种 F 近似

Rao (1951) 利用 8.5.2 节中的展开公式给出了基于 β 分布的 $U_{p,m,n}$ 的另一个函数的分布的展开式. 通过调整常量, 可以使首项后面的项的阶达到 m^{-4}. 因为 F 的自由度为 pm 和 $ks-r$, 所以下式是一个良好的近似

$$\frac{1-U^{1/s}}{U^{1/s}}\cdot\frac{ks-r}{pm}, \tag{44}$$

其中

$$s=\sqrt{\frac{p^2m^2-4}{p^2+m^2-5}},\quad r=\frac{pm}{2}-1, \tag{45}$$

k 为 $n-\frac{1}{2}(p-m-1)$. $p = 1, 2$, 或者 $m = 1, 2$ 时, F 分布与 8.4 节给出的相同. 若 $ks-r$ 不是整数, 可以在两个整数之间使用插值. 对于较小的 m, 该近似比 χ^2 近似要精确.

8.6　检验线性假设的其他准则

8.6.1　根的函数

到现在为止, 对于线性假设我们只讨论了似然比检验. 本节我们考虑其他的检验方法.

设 $\hat{\Sigma}_\Omega, \hat{\boldsymbol{\beta}}_{1\Omega}, \hat{\boldsymbol{\beta}}_{2\omega}$ 是基于 N 个观测样本的 $N(\boldsymbol{\beta}\boldsymbol{z}, \boldsymbol{\Sigma})$ 中的参数的估计. 这是一个充分统计量集, 我们将基于它们寻找检验方法. 就像 8.3 节一样, 若假设为 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$, 我们可以把假设变为 $\boldsymbol{\beta}_1 = \boldsymbol{0}$ (通过用 $\boldsymbol{x}_\alpha - \boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)}$ 替换 $\boldsymbol{x}_\alpha$). 另外

$$\begin{aligned}\boldsymbol{\beta}\boldsymbol{z}_\alpha &= \boldsymbol{\beta}_1\boldsymbol{z}_\alpha^{(1)}+\boldsymbol{\beta}_2\boldsymbol{z}_\alpha^{(2)} \\ &= \boldsymbol{\beta}_1(\boldsymbol{z}_\alpha^{(1)}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{z}_\alpha^{(2)})+(\boldsymbol{\beta}_2+\boldsymbol{\beta}_1\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1})\boldsymbol{z}_\alpha^{(2)} \\ &= \boldsymbol{\beta}_1\boldsymbol{z}_\alpha^{*(1)}+\boldsymbol{\beta}_2^*\boldsymbol{z}_\alpha^{(2)},\end{aligned} \tag{1}$$

其中 $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{(2)\prime} = \mathbf{0}$ 和 $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{*(1)\prime} = A_{11\cdot2}$. 则 $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{\beta}}_{1\Omega}$ 和 $\hat{\boldsymbol{\beta}}_2^* = \hat{\boldsymbol{\beta}}_{2\omega}$.

利用不变性原则来减少所需考虑的检验集. 如果做变换 $\boldsymbol{X}_\alpha^* = \boldsymbol{X}_\alpha + \boldsymbol{\Gamma} z_\alpha^{(2)}$, 原假设保持不变, 因为 $\mathrm{E}(\boldsymbol{X}_\alpha^*) = \boldsymbol{\beta}_1 z_\alpha^{*(1)} + (\boldsymbol{\beta}_2^* + \boldsymbol{\Gamma}) z_\alpha^{(2)}$ 和 $\boldsymbol{\beta}_2^* + \boldsymbol{\Gamma}$ 是不确定的. 充分统计量的唯一不变量为 $\hat{\boldsymbol{\Sigma}}$ 和 $\hat{\boldsymbol{\beta}}_1$ (因为对每一个 $\hat{\boldsymbol{\beta}}_2^*$, 存在一个 $\boldsymbol{\Gamma}$ 使它变为 0, 即 $-\hat{\boldsymbol{\beta}}_2^*$).

第二, 原假设在变换 $z_\alpha^{**(1)} = C z_\alpha^{*(1)}$ 下是不变的 (C 非奇异), 变换把 $\boldsymbol{\beta}_1$ 变成 $\boldsymbol{\beta}_1 C^{-1}$. 在该变换下 $\hat{\boldsymbol{\Sigma}}$ 和 $\hat{\boldsymbol{\beta}}_1 A_{11\cdot2} \hat{\boldsymbol{\beta}}_1'$ 是不变的, 我们把 $A_{11\cdot2}$ 看作与推断有关的信息. 然而, 这些是唯一的不变量. 考虑 $\hat{\boldsymbol{\beta}}_1$ 和 $A_{11\cdot2}$ 的一个函数, 比如 $f(\hat{\boldsymbol{\beta}}_1, A_{11\cdot2})$. 则存在一个 C^* 使它变为 $f(\hat{\boldsymbol{\beta}}_1 C^{*-1}, \boldsymbol{I})$, 进一步的正交变换使它变为 $f(\boldsymbol{T}, \boldsymbol{I})$, 其中 $t_{iv} = 0, i < v, t_{ii} \geqslant 0$. (如果把 $\boldsymbol{T}$ 的每一行看成一个 q_1 维的向量, 旋转坐标轴可以使第一个向量沿着第一个坐标轴, 第二个向量在前两个坐标轴所决定的平面内, 如此进行.) 但是 $\boldsymbol{T}$ 是 $\boldsymbol{T}\boldsymbol{T}' = \hat{\boldsymbol{\beta}}_1 A_{11\cdot2} \hat{\boldsymbol{\beta}}_1'$ 的一个函数, 即 $\boldsymbol{T}$ 的元素可以被这个方程和前面的限制条件唯一确定. 因此我们的检验依赖于 $\hat{\boldsymbol{\Sigma}}$ 和 $\hat{\boldsymbol{\beta}}_1 A_{11\cdot2} \hat{\boldsymbol{\beta}}_1'$. 设 $N\hat{\boldsymbol{\Sigma}} = \boldsymbol{G}$ 和 $\hat{\boldsymbol{\beta}}_1 A_{11\cdot2} \hat{\boldsymbol{\beta}}_1' = \boldsymbol{H}$.

第三, 当 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\beta}_2^*$ 不确定时, 把 $\boldsymbol{x}_\alpha$ 换成 $\boldsymbol{K}\boldsymbol{x}_\alpha$ 时原假设是不变的. 该变换把 $\boldsymbol{G}$ 变为 $\boldsymbol{K}\boldsymbol{G}\boldsymbol{K}'$, 把 $\boldsymbol{H}$ 变为 $\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}'$. 在该变换下, $\boldsymbol{G}$和 $\boldsymbol{H}$ 的唯一不变量为

$$|\boldsymbol{H} - l\boldsymbol{G}| = 0 \tag{2}$$

的根. 显然

$$\begin{aligned} 0 &= |\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}' - l\boldsymbol{K}\boldsymbol{G}\boldsymbol{K}'| \\ &= |\boldsymbol{K}(\boldsymbol{H} - l\boldsymbol{G})\boldsymbol{K}'| \\ &= |\boldsymbol{K}| \cdot |\boldsymbol{H} - l\boldsymbol{G}| \cdot |\boldsymbol{K}'| \end{aligned} \tag{3}$$

的根是不变的. 另一方面, 这是唯一的不变量, 对于给定的 $\boldsymbol{G}$ 和 $\boldsymbol{H}$, 存在 $\boldsymbol{K}$ 满足 $\boldsymbol{K}\boldsymbol{G}\boldsymbol{K}' = \boldsymbol{I}$ 以及

$$\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}' = \boldsymbol{L} = \begin{pmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l_p \end{pmatrix}, \tag{4}$$

其中 $l_1 \geqslant \cdots \geqslant l_p$ 是 (2) 的根. (见附录中的定理 A.2.2.)

定理 8.6.1 *设 $\boldsymbol{x}_\alpha$ 为来自于 $N(\boldsymbol{\beta}_1 z_\alpha^{*(1)} + \boldsymbol{\beta}_2^* z_\alpha^{(2)}, \boldsymbol{\Sigma})$ 的观测, 其中 $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{(2)\prime} = \mathbf{0}$ 和 $\sum_\alpha z_\alpha^{*(1)} z_\alpha^{*(1)\prime} = A_{11\cdot2}$. 在变换 $\boldsymbol{x}_\alpha^* = \boldsymbol{x}_\alpha + \boldsymbol{\Gamma} z_\alpha^{(2)}, z_\alpha^{**(1)} = C z_\alpha^{*(1)}$ 和 $\boldsymbol{x}_\alpha^* = \boldsymbol{K}\boldsymbol{x}_\alpha$ 下, 充分统计量和 $A_{11\cdot2}$ 的唯一不变的函数为 (2) 的根, 其中 $\boldsymbol{G} = N\hat{\boldsymbol{\Sigma}}$ 和 $\boldsymbol{H} = \hat{\boldsymbol{\beta}}_1 A_{11\cdot2} \hat{\boldsymbol{\beta}}_1'$.*

似然比准则是

$$U = \frac{|\boldsymbol{G}|}{|\boldsymbol{G} + \boldsymbol{H}|} = \frac{|\boldsymbol{K}\boldsymbol{G}\boldsymbol{K}'|}{|\boldsymbol{K}\boldsymbol{G}\boldsymbol{K}' + \boldsymbol{K}\boldsymbol{H}\boldsymbol{K}'|} = \frac{|\boldsymbol{I}|}{|\boldsymbol{I} + \boldsymbol{L}|} \tag{5}$$

$$=\prod_{i=1}^{p}(1+l_i)^{-1}$$

的一个函数, 它在变换下显然是不变的.

直觉上, 好的检验在根比较大时应该拒绝原假设, 因为如果 $\boldsymbol{\beta}_1$ 与 0 的差别比较大, 则 $\hat{\boldsymbol{\beta}}_1$ 应该比较大, $\boldsymbol{H}$ 也同样比较大. 推荐使用的其他的准则还有 (a) $\sum l_i$, (b) $\sum l_i/(1+l_i)$, (c) $\max l_i$, (d) $\min l_i$. 当准则超过某一给定的值时, 我们就拒绝原假设.

8.6.2 Lawley-Hotelling 迹准则

设矩阵$\boldsymbol{K}$满足 $\boldsymbol{KGK}'=\boldsymbol{I}\,[\boldsymbol{G}=\boldsymbol{K}^{-1}(\boldsymbol{K}')^{-1}$, 或$\boldsymbol{G}^{-1}=\boldsymbol{K}'\boldsymbol{K}]$, 则 (4) 成立. 从而所有根的和可以表示为

$$\begin{aligned}\sum_{i=1}^{p} l_i &= \mathrm{tr}\boldsymbol{L}=\mathrm{tr}\boldsymbol{KHK}' \\ &= \mathrm{tr}\boldsymbol{HK}'\boldsymbol{K}=\mathrm{tr}\boldsymbol{HG}^{-1}.\end{aligned} \tag{6}$$

Lawley (1938), Bartlett (1939) 和 Hotelling (1947), (1951) 推荐过该准则. 如果 (6) 大于某个依赖于 p,m 和 n 的常量, 该检验方法就拒绝原假设.

$\mathrm{tr}\boldsymbol{HG}^{-1}$ 的一般分布①不像 $U_{p,m,n}$ 的分布那么容易确定. 当 $p=2$ 时, Hotelling (1951) 得到了 $\mathrm{tr}\boldsymbol{HG}^{-1}=l_1+l_2$ 的分布的一个显式表达式. 第 13 章利用根 l_1 和 l_2 的密度得到了它略微不同的分布,

$$\Pr\{\mathrm{tr}\boldsymbol{HG}^{-1}\leqslant w\}=\boldsymbol{I}_{w/(2+w)}(m-1,n-1) \tag{7}$$

$$-\frac{\sqrt{\pi}\Gamma[\frac{1}{2}(m+n-1)]}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)}(1+w)^{-\frac{1}{2}(n-1)}\boldsymbol{I}_{w^2/(2+w)^2}\left[\frac{1}{2}(m-1),\frac{1}{2}(n-1)\right],$$

其中 $\boldsymbol{I}_x(a,b)$ 表示**不完全 β 函数**, 即 $\beta(y;a,b)$ 从 0 到 x 的积分.

Constantine (1966) 把 $\mathrm{tr}\boldsymbol{HG}^{-1}$ 的密度表示为推广的拉盖尔 (Laguerre) 多项式的无穷级数形式或带多项式的无穷级数形式; 然而, 这些级数只在 $\mathrm{tr}\boldsymbol{HG}^{-1}\leqslant 1$ 时收敛. Davis (1968) 说明了这些级数的分析连续性满足一组次数为 p 的齐次微分方程. Davis (1970a, 1970b) 利用解进行计算, 得到了附录 B 中的表格.

在原假设下, $\boldsymbol{G}$与 $\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}_\alpha'\,(n=N-q)$ 同分布, $\boldsymbol{H}$ 与 $\sum_{v=1}^{q_1}\boldsymbol{Y}_v\boldsymbol{Y}_v'$ 同分布, 其中 $\boldsymbol{Z}_\alpha$ 和 $\boldsymbol{Y}_v$ 是独立的, 且都服从 $N(\boldsymbol{0},\boldsymbol{\Sigma})$ 分布. 因为在具体的线性变换下, 根是不变的, 我们可以选择 $\boldsymbol{K}$ 使 $\boldsymbol{K\Sigma K}'=\boldsymbol{I}$, $\boldsymbol{G}^*=\boldsymbol{KGK}'[=\sum(\boldsymbol{KZ}_\alpha)(\boldsymbol{KZ}_\alpha)']$ 和 $\boldsymbol{H}^*=\boldsymbol{KHK}'$. 这就等价于在最初假设 $\boldsymbol{\Sigma}=\boldsymbol{I}$.

现在

① Lawley (1938) 声称得到了精确的分布, 但是结果是错误的.

$$\operatorname*{plim}_{N\to\infty}\frac{1}{N}\boldsymbol{G}=\operatorname*{plim}_{N\to\infty}\frac{n}{n+q}\frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha=\boldsymbol{I}. \tag{8}$$

该结果的求得是对 $\frac{1}{n}\boldsymbol{G}$ 的元素利用 (弱) 大数定律,

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{Z}_{i\alpha}\boldsymbol{Z}_{j\alpha}=\mathrm{E}(\boldsymbol{Z}_{i\alpha}\boldsymbol{Z}_{j\alpha})=\delta_{ij}. \tag{9}$$

定理 8.6.2 设当 $\boldsymbol{H}$ 与 $\sum_{v=1}^{q_1}\boldsymbol{Y}_v\boldsymbol{Y}'_v$ 同分布时, 其中 $\boldsymbol{Y}_v$ 相互独立且服从 $N(\boldsymbol{0},\boldsymbol{I})$ 分布, 函数 $f(\boldsymbol{H})$ 的不连续点集为零概率集. 则 $f(N\boldsymbol{H}\boldsymbol{G}^{-1})$ 的极限分布为 $f(\boldsymbol{H})$ 的分布.

证明 该定理是一个普通定理的直接应用; 该定理 [例如, Chernoff (1956) 中的定理 2] 为: 设 $\boldsymbol{X}_n$ 依分布收敛于 $\boldsymbol{X}$ (在后者的每一个连续点), 若依赖于 $\boldsymbol{X}$ 的分布, 函数 $g(x)$ 的不连续点组成一个零概率集, 则 $g(\boldsymbol{X}_n)$ 依分布收敛于 $g(\boldsymbol{X})$. 在本问题中, $\boldsymbol{X}_n$ 由 $\boldsymbol{H}$ 和 $\boldsymbol{G}$ 的元素组成, $\boldsymbol{X}$ 由 $\boldsymbol{H}$ 和 $\boldsymbol{I}$ 的元素组成. ■

推论 8.6.1 $N\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 或 $n\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 的极限分布是自由度为 pq_1 的 χ^2 分布.

因为

$$\mathrm{tr}\boldsymbol{H}=\sum_{i=1}^{p}h_{ii}=\sum_{i=1}^{p}\sum_{v=1}^{q_1}Y_{iv}^2, \tag{10}$$

该结论由定理 8.6.2 可得.

Ito (1956), (1960) 得到了渐近公式, Fujikoshi (1973) 推广了这些结果. 设 $w_{p,m,n}(\alpha)$ 为 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 的 α 分位数, 即

$$\Pr\{\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}\geqslant w_{p,m,n}(\alpha)\}=\alpha, \tag{11}$$

设 $\chi_k^2(\alpha)$ 是自由度为 k 的 χ^2 分布的 α 分位数. 则

$$nw_{p,m,n}(\alpha)=\chi_{pm}^2(\alpha)+\frac{1}{2n}\left[\frac{p+m+1}{pm+2}\chi_{pm}^4(\alpha)+(p-m+1)\chi_{pm}^2(\alpha)\right]+O(n^{-2}). \tag{12}$$

Ito 还给出了 n^{-2} 阶项. 见 Muirhead (1970). Davis (1970a), (1970b) 计算了近似的精确性. Ito 还得到了

$$\Pr\{n\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}\leqslant z\}=G_{pm}(z)-\frac{1}{2n}\left[\frac{p+m+1}{pm+2}z^2+(p-m+1)g_{pm}(z)\right]+O(n^{-2}), \tag{13}$$

其中 $G_k(z)=\Pr\{\chi_k^2\leqslant z\}$ 和 $g_k(z)=(\mathrm{d}/\mathrm{d}z)G_k(z)$. Pillai (1956) 推荐 $nw_{p,m,n}(\alpha)$ 的另一个近似, Pillai and Samson (1959) 给出了 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 的矩. Pillai and Young (1971) 和 Krishnaiah and Chang (1972) 计算了 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 的拉普拉斯变换并给出进行逆变换的方法. Khatri and Pillai (1966) 推荐一种基于矩的近似分布. Pillai and Young (1971) 建议使用基于前三阶矩的近似分布.

Grubbs (1954) 给出了 $p=2$, Davis (1970a) 给出了 $p=3,4$, Davis (1970b) 给出了 $p=5$, Davis (1980) 给出了 $p=6(1)10$ 时的分位数表; Pillai (1960) 给出了近似分位数表. Davis 的表在表 B.2 中给出.

8.6.3 Bartlett-Nanda-Pillai 迹准则

Bartlett (1939), Nanda (1950) 和 Pillai (1955) 推荐使用另一个准则:

$$\begin{aligned} V &= \sum_{i=1}^{p} \frac{l_i}{1+l_i} = \operatorname{tr}\boldsymbol{L}(\boldsymbol{I}+\boldsymbol{L})^{-1} \\ &= \operatorname{tr}\boldsymbol{KHK}'(\boldsymbol{KGK}'+\boldsymbol{KHK}')^{-1} \\ &= \operatorname{tr}\boldsymbol{HK}'[\boldsymbol{K}(\boldsymbol{G}+\boldsymbol{H})\boldsymbol{K}']^{-1}\boldsymbol{K} \\ &= \operatorname{tr}\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}, \end{aligned} \tag{14}$$

其中, 就像前面定义的, $\boldsymbol{K}$ 是使得 $\boldsymbol{KGK}' = \boldsymbol{I}$ 和 (4) 成立的矩阵. 对于

$$|\boldsymbol{H} - f(\boldsymbol{G}+\boldsymbol{H})| = 0 \tag{15}$$

的根 $f_i = \frac{l_i}{(1+l_i)}, i = 1, \cdots, p$, 准则为 $\sum_{i=1}^{p} f_i$. 原则上, 在原假设下, 分布函数、密度函数和矩都可以通过根的密度 (13.2.3 节) 求得,

$$C \prod_{i=1}^{p} f_i^{\frac{1}{2}(m-p-1)} \prod_{i=1}^{p} (1-f_i)^{\frac{1}{2}(n-p-1)} \prod_{i<j} (f_i - f_j), \tag{16}$$

其中当 $1 > f_1 > \cdots > f_p > 0$ 时,

$$C = \frac{\pi^{\frac{1}{2}p^2} \Gamma_p[\frac{1}{2}(m+n)]}{\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}m)\Gamma_p(\frac{1}{2}p)}, \tag{17}$$

否则为 0. 若 $m-p$ 和 $n-p$ 是奇数, 密度是关于 $f_1, \cdots, f_p$ 的多项式. 则密度与根的和的分布函数都是多项式.

许多作者写了关于矩、拉普拉斯变换、密度以及分布函数等各种方法的文章. Nanda (1950) 得到了 $p = 2, 3, 4, m = p+1$ 时的分布. Pillai (1954), (1956), (1960) 和 Pillai and Mijares (1959) 计算了 V 的前四阶矩, 并使用基于前四阶矩的 β 分布近似该分布. Pillai and Jayachandran (1970) 把矩母函数看作行列式的加权和进行计算, 行列式的元素是不完全 Γ 函数; 他们得到了在某些特殊情况下的准确密度, 并利用它们给出了分位数表. Krishnaiah and Chang (1972) 把分布表示为一些双重积分的逆拉普拉斯变换的线性组合, 并且为了得到分布推广了该方法. Davis (1972b) 证明了该分布满足一个微分方程, 并指出了方程的解的性质. Khatri and Pillai (1968) 得到了级数形式的分布 (非原假设下). James (1964) 给出了特征函数 (原假设下). Pillai and Jayachandran (1967) 得到了 $p = 2$ 时非原假设下的分布并计算了幂函数. 更深入的讨论请参阅 Krishnaiah (1978).

我们转来讨论渐近理论. 由定理 8.6.2 可知 nV 和 NV 的极限分布是自由度为 pm 的 χ^2 分布.

设 $v_{p,m,n}(\alpha)$ 由

$$\Pr\{\mathrm{tr}\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1} \geqslant v_{p,m,n}(\alpha)\} = \alpha \tag{18}$$

定义. 则 Davis (1970a), (1970b), Fujikoshi (1973) 和 Rothenberg (1977) 给出了

$$nv_{p,m,n}(\alpha)=\chi^2_{pm}(\alpha) + \frac{1}{2n}\left[-\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + (p-m+1)\chi^2_{pm}(\alpha)\right] + O(n^{-2}). \tag{19}$$

因为可以写为 (对于似然比检验而言)

$$nu_{p,m,n}(\alpha) = \chi^2_{pm}(\alpha) + \frac{1}{2n}(p-m+1)\chi^2_{pm}(\alpha) + O(n^{-2}), \tag{20}$$

我们得到

$$nw_{p,m,n}(\alpha) = nu_{p,m,n}(\alpha) + \frac{1}{2n}\cdot\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + O(n^{-2}), \tag{21}$$

$$nv_{p,m,n}(\alpha) = nu_{p,m,n}(\alpha) + \frac{1}{2n}\cdot\frac{p+m+1}{pm+2}\chi^4_{pm}(\alpha) + O(n^{-2}). \tag{22}$$

一种渐近展开 [Muirhead (1970), Fujikoshi (1973)] 为

$$\Pr\{nV \leqslant z\} = G_{pn}(z) + \frac{pm}{4n}[(m-p-1)G_{pm}(z) + 2(p+1)G_{pm+2}(z) - (p+m+1)G_{pm+4}(z)] + O(n^{-2}). \tag{23}$$

Muirhead and Fujikoshi 给出了高阶项.

表格. Pillai (1960) 基于拟合前四阶矩的 Pearson 曲线 (即具有调整区域的 β 分布) 给出了 $p = 2(1)8$ 时 V 的 1% 和 10% 的分位数. Mijares (1964) 把表扩展到了 $p = 50$. 表 B.3 中的一些 $(n+m)V/m = \mathrm{tr}(1/m)\boldsymbol{H}\{[1/(n+m)](\boldsymbol{G}+\boldsymbol{H})\}^{-1}$ 的分位数来自于*Concise Statistical Tables*, 计算的理论基础与 Pillai 的相同. Schuurman, Krisnaiah and Chattopodhyay (1975) 给出了 $p = 2(1)5$ 时 V 的准确的分位数; 在他们的技术报告 (ARL 73-0008) 中可以找到一个更广泛的表格. 通过把一些值与*Concise Statistical Tables* (附录 B) 中相应的值对比可以看到, 它们之间的差最大为 0.003.

8.6.4 Roy 最大根准则

$\boldsymbol{HG}^{-1}$ 的任一个特征根都可以用来作为检验准则. Roy (1953) 基于他的并交原则, 建议用 $\boldsymbol{HG}^{-1}$ 的最大特征根 l_1 进行检验. 如果 l_1 大于某一特定的值, 或 $f_1 = l_1/(1+l_1) = R$ 大于某个值 $r_{p,m,n}(\alpha)$, 其中 $r_{p,m,n}(\alpha)$ 满足

$$\Pr\{R \geqslant r_{p,m,n}(\alpha)\} = \alpha, \tag{24}$$

该检验方法拒绝原假设.

在原假设下, 根 $f_1, \cdots, f_p\,(p \leqslant m)$ 的密度由 (16) 给出. 把联合密度函数在区域 $0 \leqslant f_p \leqslant \cdots \leqslant f_1 \leqslant f^*$ 上积分就得到 $R = f_1$ 的分布函数 $\Pr\{f_1 \leqslant f^*\}$. 若 $m-p$

和 $n-p$ 都是奇数, $f_1, \cdots, f_p$ 的密度是一个多项式. 这时 f_1 的分布函数是一个关于 f^* 的多项式, f_1 的密度也是一个多项式.

Roy [(1945), (1957), 附录 9] 给出了一种能够求解形式为单变量 β 密度和 β 分布函数乘积的线性组合的分布函数的积分方法. $p=2$ 时 f_1 的分布函数为

$$\begin{aligned}\Pr\{f_1 \leqslant f\} = & I_f(m-1, n-1) \\ & -\frac{\sqrt{\pi}\Gamma[\frac{1}{2}(m+n-1)]}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} f^{\frac{1}{2}(m-1)}(1-f)^{\frac{1}{2}(n-1)} I_f\left[\frac{1}{2}(m-1), \frac{1}{2}(n-1)\right].\end{aligned} \tag{25}$$

该结果的求解过程在 13.5 节给出. Roy (1957) 中的第 8 章还给出了 $p=3$ 和 $p=4$ 时的分布函数.

由定理 8.6.2 知 $n\boldsymbol{H}\boldsymbol{G}^{-1}, N\boldsymbol{H}\boldsymbol{G}^{-1}, n\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}$ 或 $N\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}$ 的最大特征根的极限分布为 $\boldsymbol{H}$ 的最大特征根的分布, 即 $W(\boldsymbol{I}, m)$. $\boldsymbol{H}$ 的根的密度在 13.3 节给出. 原则上, 最大特征根的边际密度可以通过对联合密度积分而得, 但是实际上, 那个积分要比对 $\boldsymbol{H}\boldsymbol{G}^{-1}$ 或 $\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}$ 的根的密度函数进行积分复杂得多.

关于这方面的文献太多, 在这里很难一一总结. Nanda (1948) 得到了 $p=2,3,4,5$ 时的分布. Pillai (1954), (1956), (1965), (1967) 在原假设下研究了该分布. Sugiyama and Fukutomi (1966) 和 Sugiyama (1967) 得到了其他一些结果. Pillai (1967) 得到了一个合适的分布, 该分布为不完全 β 函数的线性组合. Davis (1972a) 证明了单个有序根的密度函数满足一个微分方程, 并得到了它的递推关系. Hayakawa (1967), Khatri and Pillai (1968), Pillai and Sugiyama (1969) 和 Khatri (1972) 讨论了非中心的情况. 更详细的讨论请参阅 Krishnaiah (1978).

表格. Nanda (1951) 和 Foster and Rees (1957) 计算了 $p=2$ 时, Foster (1957) 计算了 $p=3$ 时, Foster (1958) 计算了 $p=4$ 时, Pillai (1960) 使用逼近的方法计算了 $p=2(1)6$ 时的百分位点. [也可以参阅 Pillai (1956), (1960), (1964), (1965), (1967).] Heck (1960) 给出了 $p=2(1)6$ 时的分位数表. 表 B.4 中 nl_1/m 的分位数来自于*Concise Statistical Tables*, 该表数据基于 Pillai (1967) 的近似结果.

8.6.5 功效的对比

前面给出的四个检验方法主要是基于 Wilks 的 U, Lawley-Hotelling 的 W, Bartlett-Nanda-Pillai 的 V 和 Roy 的 R. 为了说明如何从这四种检验中进行选择, 我们来比较它们的功效函数. Rothenberg 根据非原假设情况下分布的渐近展开式对前三种检验方法进行了比较.

设 $\nu_1^N, \cdots, \nu_p^N$ 为

$$|(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot 2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^*)' - \nu\boldsymbol{\Sigma}| = 0 \tag{26}$$

的根.

$$\operatorname{tr}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1^*)'\boldsymbol{\Sigma}^{-1} \tag{27}$$

服从自由度为 pm 非中心参数为 $\sum_{i=1}^p \nu_i^N$ 的 χ^2 分布. 假设 $N \to \infty$ 且 $\boldsymbol{A}_{11\cdot2}$ 无界, 非中心化参数无限地增长且功效趋近于 1. 考虑一系列备择假设, 以便不同的检验具有不同的功效, 可以得到更多的信息. 设 $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^N$ 是一个使得 $N \to \infty$ 时 $(\boldsymbol{\beta}_1^N - \boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\boldsymbol{\beta}_1^N - \boldsymbol{\beta}_1^*)'$ 接近于一个极限的矩阵序列, 那么 $\nu_1^N, \cdots, \nu_p^N$ 趋近于各自的极限值 $\nu_1, \cdots, \nu_p$. 这时 $N\operatorname{tr}\boldsymbol{H}\boldsymbol{G}^{-1}, n\operatorname{tr}\boldsymbol{H}\boldsymbol{G}^{-1}, N\operatorname{tr}\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}$ 或 $n\operatorname{tr}\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}$ 的极限分布是自由度为 pm 非中心参数为 $\sum_{i=1}^p \nu_i$ 的 χ^2 分布. $-N\ln U$ 和 $-n\ln U$ 的情况类似可得.

Rothenberg (1977) 基于以上的条件得到了

$$\begin{aligned}\Pr\{U \leqslant u_{p,m,n}(\alpha)\} = {}& 1 - G_{pm}\left[\chi_{pm}^2(\alpha)\middle|\sum_{i=1}^p \nu_i\right] \\ & -\frac{1}{2n}\left\{(p+m+1)\sum_{i=1}^p \nu_i g_{pm+4}[\chi_{pm}^2(\alpha)]\right. \\ & \left.+\sum_{i=1}^p \nu_i^2 g_{pm+6}[\chi_{pm}^2(\alpha)]\right\} + o\left(\frac{1}{n}\right),\end{aligned} \tag{28}$$

$$\begin{aligned}&\Pr\{\operatorname{tr}\boldsymbol{H}\boldsymbol{G}^{-1} \geqslant w_{p,m,n}(\alpha)\} \\ &= 1 - G_{pm}\left[\chi_{pm}^2(\alpha)\middle|\sum_{i=1}^p \nu_i\right] \\ &\quad -\frac{1}{2n}\left\{(p+m+1)\sum_{i=1}^p \nu_i g_{pm+4}[\chi_{pm}^2(\alpha)]\right. \\ &\quad +\sum_{i=1}^p \nu_i^2 g_{pm+6}[\chi_{pm}^2(\alpha)] \\ &\quad \left.-\left[\sum_{i=1}^p \nu_i^2 - \frac{p+m+1}{pm+2}\left(\sum_{i=1}^p \nu_i\right)^2\right] g_{pm+8}[\chi_{pm}^2(\alpha)]\right\} + o\left(\frac{1}{n}\right),\end{aligned} \tag{29}$$

$$\begin{aligned}&\Pr\{\operatorname{tr}\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1} \geqslant v_{p,m,n}(\alpha)\} \\ &= 1 - G_{pm}\left[\chi_{pm}^2(\alpha)\middle|\sum_{i=1}^p \nu_i\right] \\ &\quad -\frac{1}{2n}\left\{(p+m+1)\sum_{i=1}^p \nu_i g_{pm+4}[\chi_{pm}^2(\alpha)]\right. \\ &\quad +\sum_{i=1}^p \nu_i^2 g_{pm+6}[\chi_{pm}^2(\alpha)] \\ &\quad \left.+\left[\sum_{i=1}^p \nu_i^2 - \frac{p+m+1}{pm+2}\left(\sum_{i=1}^p \nu_i\right)^2\right] g_{pm+8}[\chi_{pm}^2(\alpha)]\right\} + o\left(\frac{1}{n}\right)\end{aligned} \tag{30}$$

其中 $G_f(x|y)$ 表示自由度为 f 非中心参数为 y 的非中心 χ^2 分布. $g_f(x)$ 表示自由度为 f 的 (中心)χ^2 密度. 首项为非中心的 χ^2 分布, 三个检验的功效函数都服从这个规律. 两个迹检验的功效函数与似然比检验的功效函数不同, 用 $\pm g_{pm+8}[\chi^2_{pm}(\alpha)]/(2n)$ 乘以

$$\sum_{i=1}^{p}\nu_i^2-\frac{p+m+1}{pm+2}\left(\sum_{i=1}^{p}\nu_i\right)^2=\sum_{i=1}^{p}(\nu_i-\bar{\nu})^2-\frac{p(p-1)(p+2)}{pm+2}\bar{\nu}^2, \tag{31}$$

其中 $\bar{\nu}=\sum_{i=1}^{p}\nu_i/p$. 当

$$\frac{\sigma_\nu}{\bar{\nu}}>\sqrt{\frac{(p-1)(p+2)}{pm+2}} \tag{32}$$

时上式是正的, 其中 $\sigma_\nu^2=\sum_{i=1}^{p}(\nu_i-\bar{\nu})^2/p$ 为 $\nu_1,\cdots,\nu_p$ 的 (总体) 方差, (32) 的左端为变差的系数. 如果 ν_i 是在 (32) 成立意义下的相关变量, Lawley-Hotelling 迹检验的功效大于似然比检验的功效, 进而大于 Bartlett-Nanda-Pillai 迹检验的功效 (到 $1/n$ 阶). 若 (32) 中的不等号反号, 则功效的大小顺序也反向.

对于固定的 $\nu_1,\cdots,\nu_p$, 功效之间的差别随着 n 的增大而减小. (然而, 这种比较意义不大, 因为随着 n 增大, $(\boldsymbol{\beta}_1^N-\boldsymbol{\beta}_1^*)$ 会减小, 而 $\boldsymbol{Z}'\boldsymbol{Z}$ 会增大.)

已经有大量的关于这些检验的数值比较. Schatzoff (1966b) 和 Olson (1974) 应用了蒙特卡罗方法进行比较; Mikhail (1965), Pillai and Jayachandran (1967) 和 Lee (1971a) 利用了分布的渐近展开进行比较. 他们的结论与 Rothenberg 的结论一致. 在这三种方法中, 当根大体相等时 Bartlett-Nanda-Pillai 迹检验是最优选择, 当根明显不同时 Lawley-Hotelling 迹检验更有优势. 在极大化极小意义下 Wilks 的似然比检验似乎是第二选择.

就像 8.6.4 节说明的那样, Roy 最大根在原假设下的极限分布不是 χ^2 分布, 在一系列备择假设下也不是非中心的 χ^2 分布. 因此 Rothenberg 的比较不能推广到这种情况. 事实上, 非原假设下的分布是非常难计算的. 然而 Schatzoff (1966b) 和 Olson (1974) 的蒙特卡罗结果很明确. 如果备择假设是一维的, 即 $\nu_2=\cdots=\nu_p=0$, 最大根检验具有最大的功效. 反之, 即若备择假设不是一维的, 最大根检验是很差的.

这些检验方法是稳健的. 在原假设下, 只要分布满足一定的条件, 例如四阶矩有界, $(\hat{\boldsymbol{\beta}}_1-\boldsymbol{\beta}_1^*)$ 正态化后的极限分布就是正态分布, 其均值为 $\boldsymbol{0}$, 协方差阵与 $\boldsymbol{X}$ 为正态时相同. 这时 $\hat{\boldsymbol{\Sigma}}_\Omega=(1/N)\boldsymbol{G}$ 以概率 1 收敛. 每种准则正态化后的极限分布与 $\boldsymbol{X}$ 为正态时相同. Olson (1974) 在偏离协方差齐性和偏离正态性的条件下研究了稳健性. 他得出的结论是两个迹检验和似然比检验相对稳健, 最大根检验最不稳健. 也可以参阅 Pillai and Hsu (1979).

Berndt and Savin (1977) 给出了 (见习题 8.19)

$$\operatorname{tr}\boldsymbol{H}(\boldsymbol{H}+\boldsymbol{G})^{-1}\leqslant\ln U^{-1}\leqslant\operatorname{tr}\boldsymbol{H}\boldsymbol{G}^{-1}. \tag{33}$$

如果利用 χ^2 分位数, 可能会拒绝较大的准则而接受较小的准则.

8.7 关于回归系数矩阵和置信区域的假设检验

8.7.1 假设检验

设有一组观测向量集 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 和一组固定的向量 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_N$, 其中 $\boldsymbol{x}_\alpha$ 是来自于 $N(\boldsymbol{\beta}\boldsymbol{z}_\alpha,\boldsymbol{\Sigma})$ 的观测. 设 $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$ 和 $\boldsymbol{z}'_\alpha=(\boldsymbol{z}_\alpha^{(1)'},\boldsymbol{z}_\alpha^{(2)'})$, 其中 $\boldsymbol{\beta}_1$ 和 $\boldsymbol{z}_\alpha^{(1)'}$ 有 $q_1\,(=q-q_2)$ 列. 原假设为

$$H:\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*,\tag{1}$$

其中 $\boldsymbol{\beta}_1^*$ 是一个给定的矩阵. 设所期望的显著性水平为 α. 一种检验方法就是计算

$$U=\frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\omega|},\tag{2}$$

然后与 $U_{p,q_1,n}$ 分布的 α 分位数 $u_{p,q_1,n}(\alpha)$ 进行比较. $p=2,\cdots,10$ 和 m 为偶数时, 可以使用附录 B 中的表 1. $m=2,\cdots,10$ 和 p 为偶数, 将 m 与 p 互换就可以使用相同的表. (表中的 M 保持不变.) p 和 m 都为奇数时, p 或 m 在偶数间的插值多数情形下能够给出比较精确的值. 当 n 适当大时, 可以利用渐近理论. 一个等价的方法就是计算 $\Pr\{U_{p,m,n}\leqslant U\}$, 如果它小于 α 就拒绝原假设.

还可以使用 Lawley-Hotelling 迹准则

$$\begin{aligned}W&=\operatorname{tr}(N\hat{\boldsymbol{\Sigma}}_\omega-N\hat{\boldsymbol{\Sigma}}_\Omega)(N\hat{\boldsymbol{\Sigma}}_\Omega)^{-1}\\&=\operatorname{tr}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'(N\hat{\boldsymbol{\Sigma}}_\Omega)^{-1},\end{aligned}\tag{3}$$

Pillai 迹准则

$$\begin{aligned}V&=\operatorname{tr}(N\hat{\boldsymbol{\Sigma}}_\omega-N\hat{\boldsymbol{\Sigma}}_\Omega)(N\hat{\boldsymbol{\Sigma}}_\omega)^{-1}\\&=\operatorname{tr}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'(N\hat{\boldsymbol{\Sigma}}_\omega)^{-1},\end{aligned}\tag{4}$$

或者 Roy 最大根准则 R, 其中 R 为

$$|N\hat{\boldsymbol{\Sigma}}_\omega-N\hat{\boldsymbol{\Sigma}}_\Omega-rN\hat{\boldsymbol{\Sigma}}_\Omega|=|(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'-rN\hat{\boldsymbol{\Sigma}}_\Omega|=0\tag{5}$$

的最大根. 这些准则可以参阅附录 B 中对应的表.

这里概述一种计算准则的方法. 如果设 $\boldsymbol{y}_\alpha=\boldsymbol{x}_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{z}_\alpha^{(1)}$, 则 $\boldsymbol{y}_\alpha$ 可以看成是来自于 $N(\boldsymbol{\Delta}\boldsymbol{z}_\alpha,\boldsymbol{\Sigma})$ 的观测, 其中 $\boldsymbol{\Delta}=(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)=(\boldsymbol{\beta}_1-\boldsymbol{\beta}_1^*\ \boldsymbol{\beta}_2)$. 则原假设为 $H:\boldsymbol{\Delta}_1=\boldsymbol{0}$, 且

$$\sum\boldsymbol{y}_\alpha\boldsymbol{y}'_\alpha=\sum\boldsymbol{x}_\alpha\boldsymbol{x}'_\alpha-\boldsymbol{\beta}_1^*\boldsymbol{C}'_1-\boldsymbol{C}_1\boldsymbol{\beta}_1^{*'}+\boldsymbol{\beta}_1^*\boldsymbol{A}_{11}\boldsymbol{\beta}_1^{*'},\tag{6}$$

$$\sum\boldsymbol{y}_\alpha\boldsymbol{z}'_\alpha=\boldsymbol{C}-\boldsymbol{\beta}_1^*(\boldsymbol{A}_{11}\ \boldsymbol{A}_{12}).\tag{7}$$

因此检验假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 的问题, 等价于检验假设 $\boldsymbol{\Delta}_1=\mathbf{0}$, 其中 $\mathrm{E}(\boldsymbol{y}_\alpha)=\boldsymbol{\Delta}\boldsymbol{z}_\alpha$. 因此我们可以把原问题转化为检验假设 $\boldsymbol{\beta}_1=\mathbf{0}$. 则 $N\hat{\boldsymbol{\Sigma}}_\omega=\sum\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'$, $N\hat{\boldsymbol{\Sigma}}_\Omega=\sum\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{A}\hat{\boldsymbol{\beta}}_\Omega'$. 8.2.2 节给出了 $\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{A}\hat{\boldsymbol{\beta}}_\Omega'$ 的计算方法, 因此可得 $N\hat{\boldsymbol{\Sigma}}_\Omega$. 同样可得 $\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'$. 根据下式

$$\begin{pmatrix}\boldsymbol{A}_{22} & \boldsymbol{A}_{21}\\ \boldsymbol{A}_{12} & \boldsymbol{A}_{11}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_{2\Omega}'\\ \hat{\boldsymbol{\beta}}_{1\Omega}'\end{pmatrix}=\begin{pmatrix}\boldsymbol{C}_2'\\ \boldsymbol{C}_1'\end{pmatrix}, \tag{8}$$

$\boldsymbol{A}^*$ 和 $\boldsymbol{A}^{**}$ 的前 q_2 行和前 q_2 列与对

$$\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'=\boldsymbol{C}_2' \tag{9}$$

的左边使用向前解法得到的结果是相同的, $\bar{\boldsymbol{C}}^*$ 和 $\bar{\boldsymbol{C}}^{**}$ 的前 q_2 行等于对 (9) 式右边使用向前解法得到的结果. 因此 $\hat{\boldsymbol{\beta}}_{2\omega}\boldsymbol{A}_{22}\hat{\boldsymbol{\beta}}_{2\omega}'=\bar{\boldsymbol{C}}_2^*\bar{\boldsymbol{C}}_2^{**'}$, 其中 $\bar{\boldsymbol{C}}^{*'}=(\bar{\boldsymbol{C}}_2^{*'},\bar{\boldsymbol{C}}_1^{*'})$, $\bar{\boldsymbol{C}}^{**'}=(\bar{\boldsymbol{C}}_2^{**'},\bar{\boldsymbol{C}}_1^{**'})$.

该方法暗含了一种计算行列式的方法. 在附录 A.5 中给出向前解法的结果 $\boldsymbol{F}\boldsymbol{A}=\boldsymbol{A}^*$, 因此 $|\boldsymbol{F}||\boldsymbol{A}|=|\boldsymbol{A}^*|$. 因为三角矩阵的行列式为其对角元的乘积, 所以 $|\boldsymbol{F}|=1$ 并且 $|\boldsymbol{A}|=|\boldsymbol{A}^*|=\prod_{i=1}^q a_{ii}^*$. 只要$\boldsymbol{A}$是正定矩阵 ($\boldsymbol{F}$做恰当的调整) 该结论就成立, 并且可以用来计算 $|N\hat{\boldsymbol{\Sigma}}_\Omega|$ 和 $|N\hat{\boldsymbol{\Sigma}}_\omega|$.

8.7.2 基于 U 的置信区域

前面讨论了给定 $\boldsymbol{\beta}_1^*$ 时的假设检验 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 的问题. 通常情况下, 我们可以由检验族得到 $\boldsymbol{\beta}_1$ 的一个置信区域. 由前面的理论可知, 所抽样本满足

$$\frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\Omega+(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)\boldsymbol{A}_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)'|}\geqslant u_{p,q_1,n}(\alpha) \tag{10}$$

的概率为 $1-\alpha$. 所以如果假设 $\boldsymbol{\beta}_1$ 的置信区域满足

$$\frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\Omega+(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1)\boldsymbol{A}_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1)'|}\geqslant u_{p,q_1,n}(\alpha), \tag{11}$$

其中 (11) 是 $\boldsymbol{\beta}_1=\bar{\boldsymbol{\beta}}_1$ 时的不等式, 则所抽样本使以上假设成立的概率为 $1-\alpha$.

定理 8.7.1 *在 $\bar{\boldsymbol{\beta}}_1$ 空间中的区域 (11) 是 $\boldsymbol{\beta}_1$ 的置信系数为 $1-\alpha$ 的置信区域.*

一般情况下, 满足 (11) 的 $\bar{\boldsymbol{\beta}}_1$ 的集合是很难确定的. 但是该不等式可以用来确定 trial (测试) 矩阵是否包含在该区域内.

8.7.3 基于 Lawley-Hotelling 迹的联合置信区间

每个检验方法都暗含着一个置信区域集. Lawley-Hotelling 迹准则可以用来建立 $\boldsymbol{\beta}_1$ 的元素的线性组合的联合置信区间. 一个置信系数为 $1-\alpha$ 的置信区域为

$$\operatorname{tr}(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1')\boldsymbol{A}_{11\cdot 2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\bar{\boldsymbol{\beta}}_1')'(N\hat{\boldsymbol{\Sigma}}_\Omega)^{-1}\leqslant w_{p,m,n}(\alpha). \tag{12}$$

为了得到置信界, 我们推广引理 5.3.2.

引理 8.7.1 对于正定矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{G}$

$$|\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{Y}| \leqslant \sqrt{\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi}}\sqrt{\mathrm{tr}\boldsymbol{A}\boldsymbol{Y}'\boldsymbol{G}^{-1}\boldsymbol{Y}}. \tag{13}$$

证明 设 $b = \mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{Y}/\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi}$. 则

$$\begin{aligned}0 &\leqslant \mathrm{tr}\boldsymbol{A}(\boldsymbol{Y} - b\boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{A}^{-1})'\boldsymbol{G}^{-1}(\boldsymbol{Y} - b\boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{A}^{-1}) \\ &= \mathrm{tr}\boldsymbol{A}\boldsymbol{Y}'\boldsymbol{G}^{-1}\boldsymbol{Y} - b\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{Y} - b\mathrm{tr}\boldsymbol{Y}'\boldsymbol{\Phi} + b^2\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{A}^{-1} \\ &= \mathrm{tr}\boldsymbol{A}\boldsymbol{Y}'\boldsymbol{G}^{-1}\boldsymbol{Y} - \frac{(\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{Y})^2}{\mathrm{tr}\boldsymbol{A}^{-1}\boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi}},\end{aligned} \tag{14}$$

(13) 得证. ■

对于任意的 $p \times m$ 矩阵 $\boldsymbol{\Phi}$, 由 (12) 和 (13) 可得

$$\begin{aligned}&|\mathrm{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} - \mathrm{tr}\boldsymbol{\Phi}'\bar{\boldsymbol{\beta}}_1| = |\mathrm{tr}\boldsymbol{\Phi}'(\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1)| \\ &\leqslant \sqrt{\mathrm{tr}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{\Phi}'N\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}\cdot\mathrm{tr}\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1)'(N\hat{\boldsymbol{\Sigma}}_\Omega)^{-1}(\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1)} \\ &\leqslant \sqrt{\mathrm{tr}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{\Phi}'N\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)}.\end{aligned} \tag{15}$$

因此对所有置信度为 $1-\alpha$ 的 $\boldsymbol{\Phi}$, 有

$$\begin{aligned}&\mathrm{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} - \sqrt{N\mathrm{tr}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{\Phi}'\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)} \leqslant \mathrm{tr}\boldsymbol{\Phi}'\bar{\boldsymbol{\beta}}_1 \\ &\leqslant \mathrm{tr}\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega} + \sqrt{N\mathrm{tr}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{\Phi}'\hat{\boldsymbol{\Sigma}}_\Omega\boldsymbol{\Phi}}\sqrt{w_{p,m,n}(\alpha)}.\end{aligned} \tag{16}$$

对于各种 $\boldsymbol{\Phi}$, 我们利用 (16) 来探究置信区域 (12). 若对于某对 $(\boldsymbol{I},\boldsymbol{K})$ 有 $\phi_{ik} = 1$, 其他为 0, 则 (16) 给出了 β_{IK} 的一个区间. 若对于某对 $(\boldsymbol{I},\boldsymbol{K})$ 有 $\phi_{ik} = 1$, 对于某对 $(\boldsymbol{I},\boldsymbol{L})$ 有 $\phi_{il} = -1$, 其他为 0, 可得两个独立变量系数的差 $\beta_{IK} - \beta_{IL}$ 的区间. 若对于某对 $(\boldsymbol{I},\boldsymbol{K})$ 有 $\phi_{ik} = 1$, 对于某对 $(\boldsymbol{J},\boldsymbol{K})$ 有 $\phi_{jk} = -1$, 其他为 0, 可得两个相依变量系数的差 $\beta_{IK} - \beta_{JK}$ 的区间.

8.7.4 基于 Roy 最大根准则的联合置信区间

基于最大根准则的置信度为 $1-\alpha$ 的置信区域为

$$\mathrm{ch}_1(\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \bar{\boldsymbol{\beta}}_1)'(N\hat{\boldsymbol{\Sigma}}_\Omega)^{-1} \leqslant r_{p,m,n}(\alpha), \tag{17}$$

其中 $\mathrm{ch}_1(\boldsymbol{C})$ 表示 $\boldsymbol{C}$ 的最大特征根. 由 (17) 可以得到联合置信界. 由引理 5.3.2 可知, 对任意向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$,

$$\begin{aligned}\left[\boldsymbol{a}'(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)\boldsymbol{b}\right]^2 &= \left\{\left[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\right]'\boldsymbol{b}\right\}^2 \\ &\leqslant \left[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\right]'\boldsymbol{A}_{11\cdot2}\left[(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}\right]\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b} \\ &= \frac{\boldsymbol{a}'(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega} - \boldsymbol{\beta}_1)'\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}}\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}\end{aligned} \tag{18}$$

$$\leqslant \mathrm{ch}_1\left[(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)\boldsymbol{A}_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)'\boldsymbol{G}^{-1}\right]\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}$$
$$\leqslant r_{p,m,n}(\alpha)\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}$$

的概率为 $1-\alpha$, 第二个不等式由附录中的定理 A.2.4 可得. 一个关于所有线性组合 $\boldsymbol{a}'\boldsymbol{\beta}_1\boldsymbol{b}$ 的置信度为 $1-\alpha$ 的置信区间集为

$$\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b}-\sqrt{r_{p,m,n}(\alpha)\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}}\leqslant\boldsymbol{a}'\boldsymbol{\beta}_1\boldsymbol{b} \tag{19}$$
$$\leqslant\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b}+\sqrt{r_{p,m,n}(\alpha)\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}}.$$

线性组合为 $\boldsymbol{a}'\boldsymbol{\beta}_1\boldsymbol{b}=\sum_{i=1}^{p}\sum_{h=1}^{m}a_i\beta_{ih}b_h$. 若 $a_1=1, a_i=0, i\neq1$ 而且 $b_1=1, b_h=0, h\neq1$, 则线性组合仅为 β_{11}. 若 $a_1=1, a_i=0, i\neq1$ 而且 $b_1=1, b_2=-1, b_h=0, h\neq1,2$, 则线性组合为 $\beta_{11}-\beta_{12}$.

我们将这些区间与 (16) 在 $\boldsymbol{\Phi}=\boldsymbol{a}\boldsymbol{b}'$ 时进行比较, 其中 Φ 的秩为 1. 从 $\boldsymbol{\Phi}'\hat{\boldsymbol{\beta}}_{1\Omega}=\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b}$ 加上或减去的项为

$$w_{p,m,n}(\alpha)\cdot\mathrm{tr}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b}\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\boldsymbol{b}'=w_{p,m,n}(\alpha)\cdot\boldsymbol{a}'\boldsymbol{G}\boldsymbol{a}\cdot\boldsymbol{b}'\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{b} \tag{20}$$

的平方根. 因为 $w_{p,m,n}(\alpha)$ 大于 $r_{p,m,n}(\alpha)$, 该项大于从 (19) 式中 $\boldsymbol{a}'\hat{\boldsymbol{\beta}}_{1\Omega}\boldsymbol{b}$ 加上或减去的项, 其中 $w_{p,m,n}(\alpha)$ 与根的和有关而 $r_{p,m,n}(\alpha)$ 只与一个根有关. (16) 中的界对所有的 $p\times m$ 矩阵 $\boldsymbol{\Phi}$ 都成立, 而 (19) 只对秩为 1 的矩阵 $\boldsymbol{a}\boldsymbol{b}'$ 成立.

Mudholkar (1966) 基于对称度规函数, 给出了一种构造联合置信区间的一般方法. Gabriel (1969) 把置信界与联合检验方法联系起来. Wijsman (1979) 证明了在某些条件下, 基于最大根的置信集是最小的. [也可以参阅 Wijsman (1980).]

8.8 具有相同协方差阵的几个正态分布均值相等的检验

众所周知, 单变量分析中的许多假设都可以转化成关于回归系数的假设. 多元情况也是类似的. 作为一个例子我们考虑关于均值的假设: 具有相同协方差阵的 q 个正态分布的均值是相等的.

设 $\boldsymbol{y}_\alpha^{(i)}$ 为来自 $N(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma})$ 的观测, $\alpha=1,\cdots,N_i, i=1,\cdots,q$. 原假设为

$$H:\boldsymbol{\mu}^{(1)}=\cdots=\boldsymbol{\mu}^{(q)}. \tag{1}$$

为了把该问题化成本章一开始的形式, 令

$$\boldsymbol{X}=(\boldsymbol{x}_1\ \boldsymbol{x}_2\cdots\boldsymbol{x}_{N_1}\ \boldsymbol{x}_{N_1+1}\cdots\boldsymbol{x}_N)=(\boldsymbol{y}_1^{(1)}\ \boldsymbol{y}_2^{(2)}\cdots\boldsymbol{y}_{N_1}^{(1)}\ \boldsymbol{y}_1^{(1)}\cdots\boldsymbol{y}_{N_q}^{(q)}), \tag{2}$$

其中 $N=N_1+\cdots+N_q$. 令

$$\boldsymbol{Z}=(\boldsymbol{z}_1\ \boldsymbol{z}_2\cdots\boldsymbol{z}_{N_1}\ \boldsymbol{z}_{N_1+1}\cdots\boldsymbol{z}_N) \tag{3}$$

$$
= \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \end{pmatrix},
$$

即对于 $i=1,\cdots,q-1$, 当 $N_1+\cdots+N_{i-1}<\alpha\leqslant N_1+\cdots+N_i$ 时, $z_{i\alpha}=1$, 否则 $z_{i\alpha}=0$, 其中 $z_{q\alpha}=1$(对所有的 α). 设 $\boldsymbol{\beta}=(\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2)$, 其中

$$
\begin{aligned} \boldsymbol{\beta}_1 &= (\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(q)},\cdots,\boldsymbol{\mu}^{(q-1)}-\boldsymbol{\mu}^{(q)}), \\ \boldsymbol{\beta}_2 &= \boldsymbol{\mu}^{(q)}. \end{aligned} \tag{4}
$$

则 $\boldsymbol{x}_\alpha$ 是来自 $N(\boldsymbol{\beta} z_\alpha,\boldsymbol{\Sigma})$ 的观测, 原假设为 $\boldsymbol{\beta}_1=\boldsymbol{0}$. 因此我们可以利用前面的理论找到检验该假设的准则.

因为

$$
\boldsymbol{A}=\sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'=\begin{pmatrix} N_1 & 0 & \cdots & 0 & N_1 \\ 0 & N_2 & \cdots & 0 & N_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & N_{q-1} & N_{q-1} \\ N_1 & N_2 & \cdots & N_{q-1} & N \end{pmatrix}, \tag{5}
$$

$$
\boldsymbol{C}=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{z}_\alpha'=\left(\sum_\alpha \boldsymbol{y}_\alpha^{(1)} \quad \sum_\alpha \boldsymbol{y}_\alpha^{(2)}\cdots\sum_\alpha \boldsymbol{y}_\alpha^{(q-1)} \quad \sum_{i,\alpha}\boldsymbol{y}_\alpha^{(i)}\right). \tag{6}
$$

这里, $\boldsymbol{A}_{22}=N$ 并且 $\boldsymbol{C}_2=\sum_{i,\alpha}\boldsymbol{y}_\alpha^{(i)}$. 所以 $\hat{\boldsymbol{\beta}}_{2\omega}=\sum_{i,\alpha}\boldsymbol{y}_\alpha^{(i)}\cdot(1/N)=\bar{\boldsymbol{y}}$,

$$
\begin{aligned} N\hat{\boldsymbol{\Sigma}}_\omega &= \sum_\alpha \boldsymbol{x}_\alpha\boldsymbol{x}_\alpha' - \bar{\boldsymbol{y}}N\bar{\boldsymbol{y}}' \\ &= \sum_{i,\alpha}\boldsymbol{y}_\alpha^{(i)}\boldsymbol{y}_\alpha^{(i)\prime} - N\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}' \\ &= \sum_{i,\alpha}(\boldsymbol{y}_\alpha^{(i)}-\bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha^{(i)}-\bar{\boldsymbol{y}})'. \end{aligned} \tag{7}
$$

对于 $\hat{\boldsymbol{\Sigma}}_\Omega$, 利用公式 $N\hat{\boldsymbol{\Sigma}}_\Omega=\sum\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\hat{\boldsymbol{\beta}}_\Omega\boldsymbol{A}\hat{\boldsymbol{\beta}}_\Omega'=\sum\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{C}'$. 令

$$
\boldsymbol{D}=\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & \cdots & -1 & 1 \end{pmatrix}, \tag{8}
$$

则

$$D^{-1}=\begin{pmatrix}1&0&\cdots&0&0\\0&1&\cdots&0&0\\\vdots&\vdots&&\vdots&\vdots\\0&0&\cdots&1&0\\1&1&\cdots&1&1\end{pmatrix}. \tag{9}$$

所以

$$\begin{aligned}CA^{-1}C'&=CD'D^{-1'}A^{-1}D^{-1}DC' \\ &=CD'(DAD')^{-1}DC' \\ &=\left(\sum_\alpha y_\alpha^{(1)}\cdots\sum_\alpha y_\alpha^{(q)}\right)\begin{pmatrix}N_1&0&\cdots&0\\0&N_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&N_q\end{pmatrix}^{-1}\begin{pmatrix}\sum_\alpha y_\alpha^{(1)'}\\\vdots\\\sum_\alpha y_\alpha^{(q)'}\end{pmatrix} \\ &=\sum_i\left(\sum_\alpha y_\alpha^{(i)}\frac{1}{N_i}\sum_\gamma y_\gamma^{(i)'}\right) \\ &=\sum_i N_i\bar{y}^{(i)}\bar{y}^{(i)'},\end{aligned} \tag{10}$$

其中 $\bar{y}^{(i)}=(1/N_i)\sum_\alpha y_\alpha^{(i)}$. 因此

$$\begin{aligned}N\hat{\Sigma}_\Omega&=\sum_{i,\alpha}y_\alpha^{(i)}y_\alpha^{(i)'}-\sum_i N_i\bar{y}^{(i)}\bar{y}^{(i)'} \\ &=\sum_{i,\alpha}(y_\alpha^{(i)}-\bar{y}^{(i)})(y_\alpha^{(i)}-\bar{y}^{(i)})'.\end{aligned} \tag{11}$$

并且当 $\mu^{(1)}=\cdots=\mu^{(q)}$ 时, $\hat{\Sigma}_\omega$ 是 Σ 的估计, $\hat{\Sigma}_\Omega$ 是基于单个样本而得的 Σ 的估计的加权平均.

若原假设为真, $|N\hat{\Sigma}_\Omega|/|N\hat{\Sigma}_\omega|$ 的分布为 $U_{p,q-1,n}$, 其中 $n=N-q$. 因此显著性水平为 α 的拒绝区域为

$$\lambda=\frac{|N\hat{\Sigma}_\Omega|}{|N\hat{\Sigma}_\omega|}<u_{p,q-1,n}(\alpha). \tag{12}$$

(12) 式的左边为 8.3 节的 (11) 式, 且得到 8.4 节 (4) 和 (5) 隐含的结果

$$\begin{aligned}N\hat{\Sigma}_\omega-N\hat{\Sigma}_\Omega&=\sum_{i,\alpha}y_\alpha^{(i)}y_\alpha^{(i)'}-N\bar{y}\bar{y}'-\left(\sum_{i,\alpha}y_\alpha^{(i)}y_\alpha^{(i)'}-\sum_i N_i\bar{y}^{(i)}\bar{y}^{(i)'}\right) \\ &=\sum N_i(\bar{y}^{(i)}-\bar{y})(\bar{y}^{(i)}-\bar{y})'=H,\end{aligned} \tag{13}$$

这里H的分布为 $W(\Sigma,q-1)$. 易见, 当 $p=1$ 时该检验就变为普通的 F 检验

$$\frac{\sum N_i(\bar{\boldsymbol{y}}^{(i)}-\bar{\boldsymbol{y}})^2}{(\boldsymbol{y}_\alpha^{(i)}-\bar{\boldsymbol{y}}^{(i)})^2}\cdot\frac{n}{q-1}>F_{q-1,n}(\alpha). \tag{14}$$

下面给出一个分析的例子. 数据来自于 Barnard (1935) 对埃及人头盖骨的研究. $4(=q)$ 个总体分别为古埃及前王朝时的头盖骨 $(i=1)$, 埃及第六王朝到第十二王朝时的头盖骨 $(i=2)$, 第十二王朝到第十三王朝时的头盖骨 $(i=3)$, 托勒密王朝时的头盖骨 $(i=4)$. $4(=p)$ 个测量 (即 $\boldsymbol{y}_\alpha^{(i)}$ 的分量) 为最大宽度, 基槽长度, 鼻高以及颅高. 观测次数分别为 $N_1=91, N_2=162, N_3=70, N_4=75$. 数据总结如下

$$\begin{aligned}&(\bar{\boldsymbol{y}}^{(1)}\quad\bar{\boldsymbol{y}}^{(2)}\quad\bar{\boldsymbol{y}}^{(3)}\quad\bar{\boldsymbol{y}}^{(4)})\\&=\begin{pmatrix}133.582\,418 & 134.265\,432 & 134.371\,429 & 135.306\,667\\ 98.307\,692 & 96.462\,963 & 95.857\,143 & 95.040\,000\\ 50.835\,165 & 51.148\,148 & 50.100\,000 & 52.093\,333\\ 133.000\,000 & 134.882\,716 & 133.642\,857 & 131.466\,667\end{pmatrix},\end{aligned} \tag{15}$$

$$\begin{aligned}&N\hat{\boldsymbol{\Sigma}}_\Omega\\&=\begin{pmatrix}9661.997\,470 & 445.573\,301 & 1130.623\,900 & 2148.584\,210\\ 445.573\,301 & 9073.115\,027 & 1239.211\,990 & 2255.812\,722\\ 1130.623\,900 & 1239.211\,990 & 3938.320\,351 & 1271.054\,662\\ 2148.584\,210 & 2255.812\,722 & 1271.054\,662 & 8741.508\,829\end{pmatrix}.\end{aligned} \tag{16}$$

由这些数据可得

$$\begin{aligned}&N\hat{\boldsymbol{\Sigma}}_\omega\\&=\begin{pmatrix}9785.178\,098 & 214.197\,666 & 1217.929\,248 & 2019.820\,216\\ 214.197\,666 & 9559.460\,890 & 1131.716\,372 & 2381.126\,040\\ 1217.929\,248 & 1131.716\,372 & 4088.731\,856 & 1133.473\,898\\ 2019.820\,216 & 2381.126\,040 & 1133.473\,898 & 9382.242\,720\end{pmatrix}.\end{aligned} \tag{17}$$

使用似然比检验. 行列式的比为

$$U=\frac{|N\hat{\boldsymbol{\Sigma}}_\Omega|}{|N\hat{\boldsymbol{\Sigma}}_\omega|}=\frac{2.426\,905\,4\times10^5}{2.954\,447\,5\times10^5}=0.821\,434\,4. \tag{18}$$

这里 $N=398, n=394, p=4, q=4$. 因此 $k=393$. 因为 n 非常大, 所以我们可以假设 $-k\ln U_{4,3,394}$ 服从自由度为 12 的 χ^2 分布 (原假设为真时). 这里 $-k\ln U=77.30$, 因为 χ^2_{12} 分布的 1% 分位点为 26.2, 假设 $\boldsymbol{\mu}^{(1)}=\boldsymbol{\mu}^{(2)}=\boldsymbol{\mu}^{(3)}=\boldsymbol{\mu}^{(4)}$ 被拒绝.[①]

① 上面的计算是由 Bartlett (1947) 给出的.

8.9 多元方差分析

单变量方差分析可以直接推广到向量变量, 即向量平方和的分析 (例如 $\sum \boldsymbol{x}_\alpha \boldsymbol{x}'_\alpha$). 事实上, 在前面的章节中, 这种推广已用于解决涉及分类的方差分析问题.

作为一个例子, 我们来考虑二因子试验设计. 假设我们对列效应是否为零感兴趣. 我们首先复习一下标量变量的分析方法然后给出向量变量的分析方法. 设 $Y_{ij}\ (i=1,\cdots,r, j=1,\cdots,c)$ 为 rc 个随机变量集. 假设

$$\mathrm{E}(Y_{ij}) = \mu + \lambda_i + \nu_j, \quad i=1,\cdots,r, j=1,\cdots,c, \tag{1}$$

约束条件为

$$\sum_{i=1}^{r} \lambda_i = \sum_{j=1}^{c} \nu_j = 0, \tag{2}$$

Y_{ij} 独立地服从方差为 σ^2 的正态分布. 检验列效应为零就是检验

$$\nu_j = 0, \quad j=1,\cdots,c. \tag{3}$$

通过引入固定哑变量, 该问题可以转化为回归问题. 设

$$\begin{aligned} z_{00,ij} &= 1, \\ z_{k0,ij} &= 1, \quad k=i, \\ &= 0, \quad k \neq i, \\ z_{0k,ij} &= 1, \quad k=j, \\ &= 0, \quad k \neq j. \end{aligned} \tag{4}$$

则 (1) 可以写为

$$\mathrm{E}(Y_{ij}) = \mu z_{00,ij} + \sum_{k=1}^{r} \lambda_k z_{k0,ij} + \sum_{k=1}^{c} \nu_k z_{0k,ij}. \tag{5}$$

该假设就是 $z_{0k,ij}$ 的系数为零. 因为固定变量的矩阵

$$\begin{pmatrix} z_{00,11} & \cdots & z_{00,rc} \\ z_{10,11} & \cdots & z_{10,rc} \\ z_{20,11} & \cdots & z_{20,rc} \\ \vdots & & \vdots \\ z_{0c,11} & \cdots & z_{0c,rc} \end{pmatrix} \tag{6}$$

是奇异的 (例如行 $10, 20, \cdots, r0$ 的和为行 00), 所以需要详述回归理论. 然后我们会发现, 由回归理论得到的检验准则就是方差分析中常用的 F 检验.

令

$$
\begin{aligned}
Y_{\cdot\cdot} &= \frac{1}{rc}\sum_{i,j} Y_{ij}, \\
Y_{i\cdot} &= \frac{1}{c}\sum_{j} Y_{ij}, \\
Y_{\cdot j} &= \frac{1}{r}\sum_{i} Y_{ij},
\end{aligned} \tag{7}
$$

以及

$$
\begin{aligned}
a &= \sum_{i,j}(Y_{ij} - Y_{i\cdot} - Y_{\cdot j} + Y_{\cdot\cdot})^2 \\
&= \sum_{i,j} Y_{ij}^2 - c\sum_{i} Y_{i\cdot}^2 - r\sum_{j} Y_{\cdot j}^2 + rcY_{\cdot\cdot}^2, \\
b &= r\sum_{j}(Y_{\cdot j} - Y_{\cdot\cdot})^2 \\
&= r\sum_{j} Y_{\cdot j}^2 - rcY_{\cdot\cdot}^2.
\end{aligned} \tag{8}
$$

则 F 统计量为

$$
F = \frac{b}{a}\cdot\frac{(c-1)(r-1)}{c-1}. \tag{9}
$$

在原假设下它服从自由度为 $c-1$ 和 $(r-1)(c-1)$ 的 F 分布. 假设的似然比准则为

$$
\frac{a}{a+b} = \frac{1}{1+\{(c-1)/[(r-1)(c-1)]\}F} \tag{10}
$$

的 $rc/2$ 次幂.

现在来看多元方差分析. 假设 p 维随机变量集 $\boldsymbol{Y}_{ij},\ i=1,\cdots,r, j=1,\cdots,c$, 相互独立地服从期望值为 (1)、协方差阵为$\boldsymbol{\Sigma}$的正态分布, 其中$\boldsymbol{\mu}$, $\boldsymbol{\lambda}$和$\boldsymbol{\nu}$是向量. 相同的代数方法可以将该问题转化为回归问题. 我们利用 (7) 定义 $\boldsymbol{Y}_{\cdot\cdot}, \boldsymbol{Y}_{i\cdot}, \boldsymbol{Y}_{\cdot j}$ 且

$$
\begin{aligned}
\boldsymbol{A} &= \sum_{i,j}(\boldsymbol{Y}_{ij} - \boldsymbol{Y}_{i\cdot} - \boldsymbol{Y}_{\cdot j} + \boldsymbol{Y}_{\cdot\cdot})(\boldsymbol{Y}_{ij} - \boldsymbol{Y}_{i\cdot} - \boldsymbol{Y}_{\cdot j} + \boldsymbol{Y}_{\cdot\cdot})' \\
&= \sum_{i,j}\boldsymbol{Y}_{ij}\boldsymbol{Y}_{ij}' - c\sum_{i}\boldsymbol{Y}_{i\cdot}\boldsymbol{Y}_{i\cdot}' - r\sum_{j}\boldsymbol{Y}_{\cdot j}\boldsymbol{Y}_{\cdot j}' + rc\boldsymbol{Y}_{\cdot\cdot}\boldsymbol{Y}_{\cdot\cdot}', \\
\boldsymbol{B} &= r\sum_{j}(\boldsymbol{Y}_{\cdot j} - \boldsymbol{Y}_{\cdot\cdot})(\boldsymbol{Y}_{\cdot j} - \boldsymbol{Y}_{\cdot\cdot})' \\
&= r\sum_{j}\boldsymbol{Y}_{\cdot j}\boldsymbol{Y}_{\cdot j}' - rc\boldsymbol{Y}_{\cdot\cdot}\boldsymbol{Y}_{\cdot\cdot}'.
\end{aligned} \tag{11}
$$

类似于 (10) 的统计量为

$$
\frac{|\boldsymbol{A}|}{|\boldsymbol{A}+\boldsymbol{B}|}. \tag{12}
$$

在原假设下它与 8.4 节中给出的 U 同分布, $p, n=(r-1)\cdot(c-1), q_1=c-1$. 为了使$\boldsymbol{A}$为非奇异矩阵 (以概率 1), 要求 $p \leqslant (r-1)(c-1)$.

作为示例, 我们使用由 Immer, Hayes and Powers (1934) 首次给出的数据, 该数据后来又被 Fisher (1947a), Yates and Cochran (1938) 和 Tukey (1949) 应用. 观测向量的第一个分量为指定年份的大麦产量, 第二个分量为下一年的大麦产量. 列指标取遍所有大麦的品种, 行指标取遍所有的种植地点. 数据在表 8.1 中给出 [例如, 左上角的 $\substack{81\\81}$ 表示在地点 UF, 品种 M 每一年的产量都是 81], 边缘数字为和.

表　8.1

地点	品种					和
	M	S	V	T	P	
UF	81	105	120	110	98	514
	81	82	80	87	84	414
W	147	142	151	192	146	778
	100	116	112	148	108	584
M	82	77	78	131	90	458
	103	105	117	140	130	595
C	120	121	124	141	125	631
	99	62	96	126	76	459
GR	99	89	69	89	104	450
	66	50	97	62	80	355
D	87	77	79	102	96	441
	68	67	67	92	94	338
和	616	611	621	765	659	3272
	517	482	569	655	572	2795

把 (147, 100) 的平方写为

$$\begin{pmatrix} 147 \\ 100 \end{pmatrix} (147\ 100) = \begin{pmatrix} 21\ 609 & 14\ 700 \\ 14\ 700 & 10\ 000 \end{pmatrix}.$$

则

$$\sum_{i,j} \boldsymbol{Y}_{ij}\boldsymbol{Y}'_{ij} = \begin{pmatrix} 380\ 944 & 315\ 381 \\ 315\ 381 & 277\ 625 \end{pmatrix}, \tag{13}$$

$$\sum_{j} (6\boldsymbol{Y}_{\cdot j})(6\boldsymbol{Y}_{\cdot j})' = \begin{pmatrix} 2\ 157\ 924 & 1\ 844\ 346 \\ 1\ 844\ 346 & 1\ 579\ 583 \end{pmatrix}, \tag{14}$$

$$\sum_{i} (5\boldsymbol{Y}_{i\cdot})(5\boldsymbol{Y}_{i\cdot})' = \begin{pmatrix} 1\ 874\ 386 & 1\ 560\ 145 \\ 1\ 560\ 145 & 1\ 353\ 727 \end{pmatrix}, \tag{15}$$

$$(30\boldsymbol{Y}_{\cdot\cdot})(30\boldsymbol{Y}_{\cdot\cdot})' = \begin{pmatrix} 10\ 750\ 984 & 9\ 145\ 240 \\ 9\ 145\ 240 & 7\ 812\ 025 \end{pmatrix}. \tag{16}$$

所以误差平方和为

$$A = \begin{pmatrix} 3279 & 802 \\ 802 & 4017 \end{pmatrix}, \tag{17}$$

行平方和为

$$5\sum_j (\boldsymbol{Y}_{i\cdot} - \boldsymbol{Y}_{\cdot\cdot})(\boldsymbol{Y}_{i\cdot} - \boldsymbol{Y}_{\cdot\cdot})' = \begin{pmatrix} 18\,011 & 7.188 \\ 7\,188 & 10\,345 \end{pmatrix}, \tag{18}$$

列平方和为

$$\boldsymbol{B} = \begin{pmatrix} 2\,788 & 2\,550 \\ 2\,550 & 2\,863 \end{pmatrix}. \tag{19}$$

检验准则为

$$\frac{|\boldsymbol{A}|}{|\boldsymbol{A}+\boldsymbol{B}|} = \frac{\begin{vmatrix} 3\,279 & 802 \\ 802 & 4\,017 \end{vmatrix}}{\begin{vmatrix} 6\,067 & 3\,352 \\ 3\,352 & 6\,880 \end{vmatrix}} = 0.410\,7. \tag{20}$$

把该结果与 $U_{2,4,20}$ 的分位数进行比较. 利用 8.4 节的结果将

$$\frac{1-\sqrt{0.410\,7}}{\sqrt{0.410\,7}} \cdot \frac{19}{4} = 2.66,$$

与 $F_{8,38}$ 的分位数进行比较. 在 5% 的显著性水平下它是显著的. 由我们的数据可知品种之间是有差异的.

单变量方差分析中的每一个 F 检验在多元方差分析中都有对应的检验. 在单变量方差分析的线性假设模型中, 我们假设随机变量 $Y_1, \cdots, Y_N$ 的期望值是未知参数的线性组合,

$$\mathrm{E}(Y_\alpha) = \sum_g \beta_g z_{g\alpha}, \tag{21}$$

其中 β 是参数, z 是已知的系数. 假设变量 $\{Y_\alpha\}$ 具有正态分布且相互独立, 公共方差为 σ^2. 在该模型下存在一个线性组合集, 假设为 $\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha$, 其中 γ 是已知的, 使得

$$a = \sum_{i=1}^n \left(\sum_{\alpha=1}^N \gamma_{i\alpha} Y_\alpha \right)^2 = \sum_{\alpha,\beta=1}^N d_{\alpha\beta} Y_\alpha Y_\beta \tag{22}$$

服从自由度为 n 的 $\sigma^2\chi^2$ 分布. 同样存在另外一个线性组合集, 假设为 $\sum_\alpha \phi_{g\alpha} Y_\alpha$, 其中 ϕ 是已知的, 使得当原假设为真时,

$$b = \sum_{g=1}^m \left(\sum_{\alpha=1}^N \phi_{g\alpha} Y_\alpha \right)^2 = \sum_{\alpha,\beta=1}^N c_{\alpha\beta} Y_\alpha Y_\beta \tag{23}$$

服从自由度为 m 的 $\sigma^2\chi^2$ 分布, 当原假设不真时它的分布为 σ^2 乘以一个非中心的 χ^2 分布. 无论哪种情况下 a 与 b 都是分布独立的, 所以当原假设为真时

$$\frac{b}{a}\cdot\frac{n}{m}=\frac{\sum c_{\alpha\beta}Y_\alpha Y_\beta}{\sum d_{\alpha\beta}Y_\alpha Y_\beta}\cdot\frac{n}{m} \tag{24}$$

服从自由度为 m 和 n 的 F 分布. 其中原假设是某些 β 为零.

在多元方差分析中, $\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_N$ 是具有 p 个分量的向量变量. $\boldsymbol{Y}_\alpha$ 的期望值由 (21) 给出, 其中 β_g 是具有 p 个参数的向量. 我们假设 $\{\boldsymbol{Y}_\alpha\}$ 独立地服从正态分布且具有公共协方差阵 $\boldsymbol{\Sigma}$. 那么线性组合 $\sum\gamma_{i\alpha}\boldsymbol{Y}_\alpha$ 是一个向量, 而且

$$\boldsymbol{A}=\sum_{i=1}^{n}\left(\sum_{\alpha=1}^{N}\gamma_{i\alpha}\boldsymbol{Y}_\alpha\right)\left(\sum_{\alpha=1}^{N}\gamma_{i\alpha}\boldsymbol{Y}_\alpha\right)'=\sum_{\alpha,\beta=1}^{N}d_{\alpha\beta}\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\beta \tag{25}$$

的分布为 $W(\boldsymbol{\Sigma},n)$. 当原假设为真时,

$$\boldsymbol{B}=\sum_{g=1}^{m}\left(\sum_{\alpha=1}^{N}\phi_{g\alpha}\boldsymbol{Y}_\alpha\right)\left(\sum_{\alpha=1}^{N}\phi_{g\alpha}\boldsymbol{Y}_\alpha\right)'=\sum_{\alpha,\beta=1}^{N}c_{\alpha\beta}\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\beta \tag{26}$$

的分布为 $W(\boldsymbol{\Sigma},m)$, 且 $\boldsymbol{B}$ 与 $\boldsymbol{A}$ 独立. 所以

$$\frac{|\boldsymbol{A}|}{|\boldsymbol{A}+\boldsymbol{B}|}=\frac{|\sum d_{\alpha\beta}\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\beta|}{|\sum d_{\alpha\beta}\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\beta+\sum c_{\alpha\beta}\boldsymbol{Y}_\alpha\boldsymbol{Y}'_\beta|} \tag{27}$$

的分布为 $U_{p,m,n}$.

在原假设所指定的那些 β 为 0 的条件下 (在未指定的 β 上是一个等式), 由关于 a 和 b 分布的讨论可知 $\mathrm{E}(\sum_\alpha\gamma_{i\alpha}Y_\alpha)=0,\mathrm{E}(\sum_\alpha\phi_{g\alpha}Y_\alpha)=0$. 很明显该结论对于向量的情形也是成立的. 单变量时, 存在一个正交矩阵 $\boldsymbol{\Psi}=(\psi_{\alpha\beta})$, 若做变换 $Y_\beta=\sum_\alpha\psi_{\beta\alpha}Z_\alpha$, 则有

$$\begin{aligned}a&=\sum_{\alpha,\beta,\gamma,\delta}d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}Z_\gamma Z_\delta=\sum_{\alpha=1}^{n}Z_\alpha^2,\\ b&=\sum_{\alpha,\beta,\gamma,\delta}c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}Z_\gamma Z_\delta=\sum_{\alpha=n+1}^{n+m}Z_\alpha^2.\end{aligned} \tag{28}$$

由于变换是正交的, 所以 $\{Z_\alpha\}$ 独立地服从正态分布且具有公共方差 σ^2. 因为 $Z_\alpha\,(\alpha=1,\cdots,n)$ 是 $\sum_\alpha\gamma_{i\alpha}Y_\alpha$ 的线性组合, $Z_\alpha\,(\alpha=n+1,\cdots,n+m)$ 是 $\sum_{\alpha=1}\phi_{g\alpha}Y_\alpha$ 的线性组合, 所以它们的均值必定为零 (在原假设下). 因此 a/σ^2 和 b/σ^2 具有上述独立的 χ^2 分布.

在多元的情况应用变换 $\boldsymbol{Y}_\beta=\sum_\alpha\psi_{\beta\alpha}\boldsymbol{Z}_\alpha$, 其中 $\boldsymbol{Y}_\beta$ 和 $\boldsymbol{Z}_\alpha$ 是向量. 那么由 (28) 可知, 当 $\gamma=\delta\leqslant n$ 时, $\sum_{\alpha,\beta}d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}=1$, 否则为 0; 当 $n+1\leqslant\gamma=\delta\leqslant n+m$ 时, $\sum_{\alpha,\beta}c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}=1$, 否则为 0, 从而

$$\begin{aligned}\boldsymbol{A}&=\sum_{\alpha,\beta,\gamma,\delta}d_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}\boldsymbol{Z}_\gamma\boldsymbol{Z}'_\delta=\sum_{\alpha=1}^{n}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha,\\ \boldsymbol{B}&=\sum_{\alpha,\beta,\gamma,\delta}c_{\alpha\beta}\psi_{\alpha\gamma}\psi_{\beta\delta}\boldsymbol{Z}_\gamma\boldsymbol{Z}'_\delta=\sum_{\alpha=n+1}^{n+m}\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha.\end{aligned} \tag{29}$$

因为 $\boldsymbol{\Psi}$ 是正交的, 所以 $\{\boldsymbol{Z}_\alpha\}$ 独立地服从协方差阵为 $\boldsymbol{\Sigma}$ 的正态分布. 在原假设下, 由相同的讨论可得 $\mathrm{E}(\boldsymbol{Z}_\alpha)=\boldsymbol{0},\alpha=1,\cdots,n+m$. 因此 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别服从 $W(\boldsymbol{\Sigma},n)$ 和 $W(\boldsymbol{\Sigma},m)$ 分布, 且相互独立.

8.10 检验的一些最优性质

8.10.1 不变检验的容许性

本章我们考虑了线性假设的几种检验, 这些检验相对于使原假设不变的变换是不变的. 但是哪些不变检验比较好呢. 特别是当我们寻找容许的检验过程时, 即在降低第 I 类错误或第 II 类错误的概率的意义下无法进一步改进的过程. 一个好检验没有必要一定是不变的. 显然, 如果一个不变检验在所有检验类中是容许的, 那么它在不变检验类中也是容许的.

这里要讨论的一般线性假设检验是第 5 章所讨论的关于均值向量的假设检验的推广. 第 8 章的不变过程是 T^2 检验的推广. 一种证明某检验过程是否容许的方法就是给参数设置一个先验分布使贝叶斯过程就是给定的检验过程. 该方法在构造先验分布时需要一些技巧, 然而给定先验分布之后的性质就比较直观了. 习题 8.26 和习题 8.27 说明了 Barlett-Nanda-Pillai 迹准则 V 和 Wilks 的似然比准则 U 都是容许的检验. 利用该方法证明容许性的缺点就是要给每个检验过程创造一个先验分布, 一般性的定理不能涵盖所有的情况.

另外一种证明容许性的方法就是利用 Stein 定理 (定理 5.6.5), 它得到了一般的结果. 不变检验可以根据行列式方程

$$|\boldsymbol{H}-\lambda(\boldsymbol{H}+\boldsymbol{G})|=0 \tag{1}$$

的根来表述, 其中 $\boldsymbol{H}=\hat{\boldsymbol{\beta}}_1\boldsymbol{A}_{11\cdot2}\hat{\boldsymbol{\beta}}_1'=\boldsymbol{W}_1\boldsymbol{W}_1',\boldsymbol{G}=N\hat{\boldsymbol{\Sigma}}_\Omega=\boldsymbol{W}_3\boldsymbol{W}_3'$. 与多余参数 $\boldsymbol{\beta}_2$ 有关的矩阵为 $\hat{\boldsymbol{\beta}}_2$(或$\boldsymbol{W}_2$). 为方便计, 在下面的记号中我们定义了典范型. 设 $\boldsymbol{W}_1=\boldsymbol{X}\,(p\times m),\boldsymbol{W}_2=\boldsymbol{Y}\,(p\times r),\boldsymbol{W}_3=\boldsymbol{Z}\,(p\times n),\mathrm{E}(\boldsymbol{X})=\boldsymbol{\Xi},\mathrm{E}(\boldsymbol{Y})=\boldsymbol{H}$ 和 $\mathrm{E}(\boldsymbol{Z})=\boldsymbol{0}$; 列独立地服从协方差阵为$\boldsymbol{\Sigma}$的正态分布. 原假设为 $\boldsymbol{\Xi}=\boldsymbol{0}$, 备择假设为 $\boldsymbol{\Xi}\neq\boldsymbol{0}$.

通常的检验以

$$|\boldsymbol{X}\boldsymbol{X}'-\lambda(\boldsymbol{Z}\boldsymbol{Z}'+\boldsymbol{X}\boldsymbol{X}')|=|\boldsymbol{X}\boldsymbol{X}'-\lambda(\boldsymbol{U}-\boldsymbol{Y}\boldsymbol{Y}')|=0 \tag{2}$$

的 (非零) 根的形式给出, 其中 $\boldsymbol{U}=\boldsymbol{X}\boldsymbol{X}'+\boldsymbol{Y}\boldsymbol{Y}'+\boldsymbol{Z}\boldsymbol{Z}'$. 根的期望等于零, (2) 的根与 $\boldsymbol{X}'(\boldsymbol{U}-\boldsymbol{Y}\boldsymbol{Y}')^{-1}\boldsymbol{X}$ 的非零特征根相同. 设 $\boldsymbol{V}=(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U})$ 并且

$$\boldsymbol{M}(\boldsymbol{V})=\boldsymbol{X}'(\boldsymbol{U}-\boldsymbol{Y}\boldsymbol{Y}')^{-1}\boldsymbol{X}. \tag{3}$$

$\boldsymbol{M}(\boldsymbol{V})$ 的有序特征根的向量记为

$$(\lambda_1, \cdots, \lambda_m)' = \boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V})), \tag{4}$$

其中 $\lambda_1 \geqslant \cdots \geqslant \lambda_m \geqslant 0$. 因为包含零根 (当 $m \geqslant p$) 不影响顺序, 我们假设检验依赖于 $\boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V}))$.

这些检验的容许性可以通过接受区域的几何特征来表述. 令

$$\begin{aligned} R^m_{\leqslant} &= \{\boldsymbol{\lambda} \in R^m | \lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant 0\}, \\ R^m_{+} &= \{\boldsymbol{\lambda} \in R^m | \lambda_1 \geqslant 0, \cdots, \lambda_m \geqslant 0\}. \end{aligned} \tag{5}$$

若基于一组样本根接受了原假设, 那么基于一组更小的根也应该接受原假设 (图 8.2).

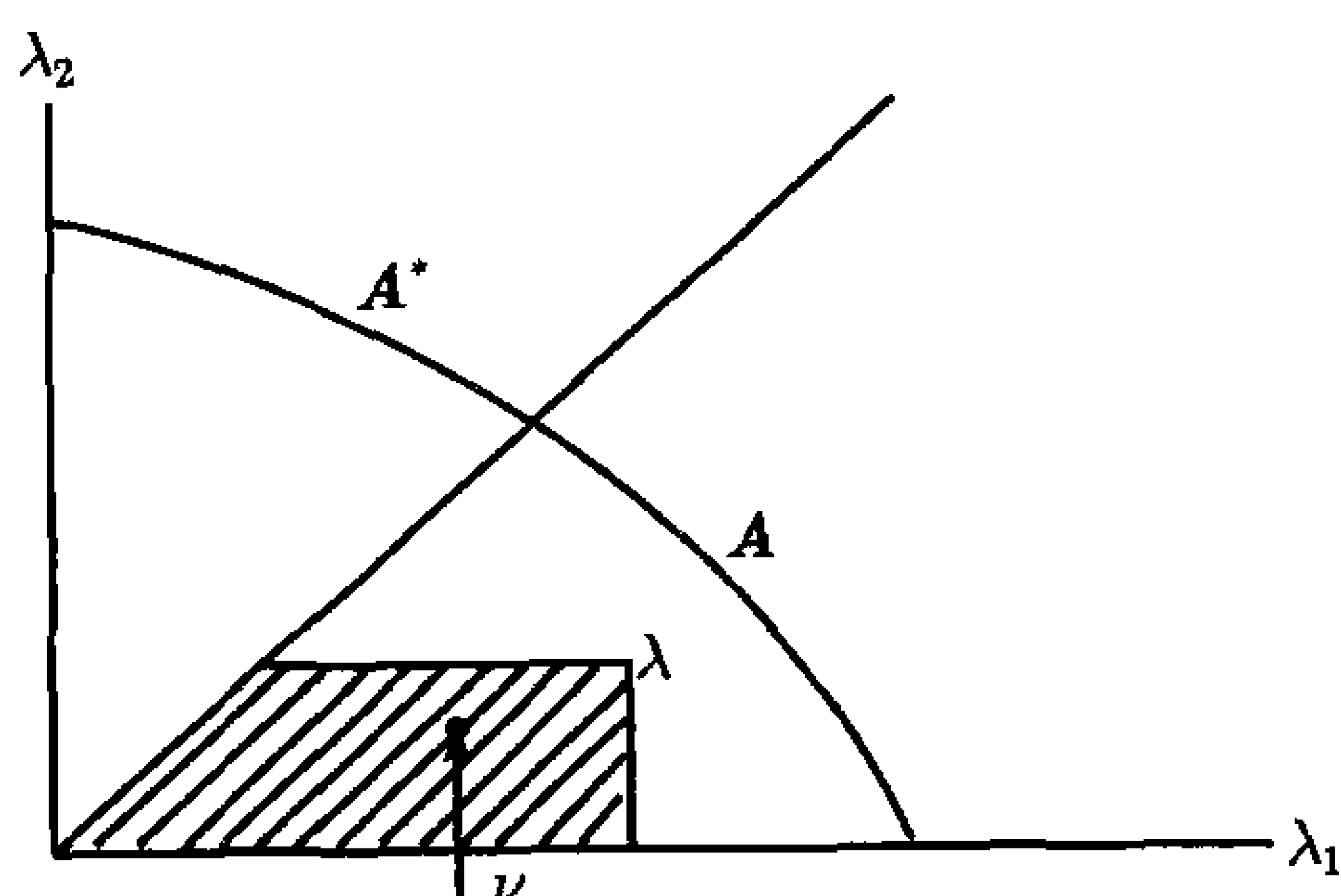


图 8.2　单调接受区域

定义 8.10.1　称一个区域 $A \subset R^m_{\leqslant}$ 是单调的, 若由 $\boldsymbol{\lambda} \in A, \boldsymbol{\nu} \in R^m_{\leqslant}$, 且 $\nu_i \leqslant \lambda_i, i = 1, \cdots, m$, 可得 $\boldsymbol{\nu} \in A$.

定义 8.10.2　对于 $A \subset R^m_{\leqslant}$, 扩展区域 A^* 定义为

$$A^* = \cup_{\pi}\{(x_{\pi(1)}, \cdots, x_{\pi(m)})' | \boldsymbol{x} \in A\} \tag{6}$$

其中 π 取遍 $(1, \cdots, m)$ 的所有置换.

下面的定理由 Schwartz(1967) 首次证明.

定理 8.10.1　若区域 $A \subset R^m_{\leqslant}$ 是单调的, 扩展区域 A^* 是闭的而且是凸的, 则 A 是一个容许检验的接受区域.

容许检验的另一个特点是优势理论.

定义 8.10.3　称一个向量 $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_m)'$ 是弱优于向量 $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_m)'$ 的, 若

$$\lambda_{[1]} \geqslant \nu_{[1]}, \lambda_{[1]} + \lambda_{[2]} \geqslant \nu_{[1]} + \nu_{[2]}, \cdots, \lambda_{[1]} + \cdots + \lambda_{[m]} \geqslant \nu_{[1]} + \cdots \nu_{[m]}, \tag{7}$$

其中 $\lambda_{[i]}$ 和 $\nu_{[i]}$ 是以非减顺序安排的坐标.

若 $\boldsymbol{\lambda}$ 弱优于 $\boldsymbol{\nu}$, 我们记为 $\boldsymbol{\lambda} \succ_w \boldsymbol{\nu}$ 或 $\boldsymbol{\nu} \prec_w \boldsymbol{\lambda}$. 若 $\boldsymbol{\lambda}, \boldsymbol{\nu} \in R^m_{\leqslant}$, 那么 $\boldsymbol{\lambda} \succ_w \boldsymbol{\nu}$ 意味着

$$\lambda_1 \geqslant \nu_1, \lambda_1 + \lambda_2 \geqslant \nu_1 + \nu_2, \cdots, \lambda_1 + \cdots + \lambda_m \geqslant \nu_1 + \cdots \nu_m. \tag{8}$$

将 (7) 中的最后一个不等式换为等式, 我们称 $\boldsymbol{\lambda}$ 优于 $\boldsymbol{\nu}$, 记为 $\boldsymbol{\lambda} \succ \boldsymbol{\nu}$ 或 $\boldsymbol{\nu} \prec \boldsymbol{\lambda}$. 优势理论和相关不等式在 Marshall and Olkin(1979) 中有详细的叙述.

定义 8.10.4 区域 $A \subset R^m_<$ 称为以优势单调的, 若由 $\boldsymbol{\lambda} \in A, \boldsymbol{\nu} \in R^m_<, \boldsymbol{\nu} \prec_w \boldsymbol{\lambda}$, 可得 $\boldsymbol{\nu} \in A$.(见图 8.3.)

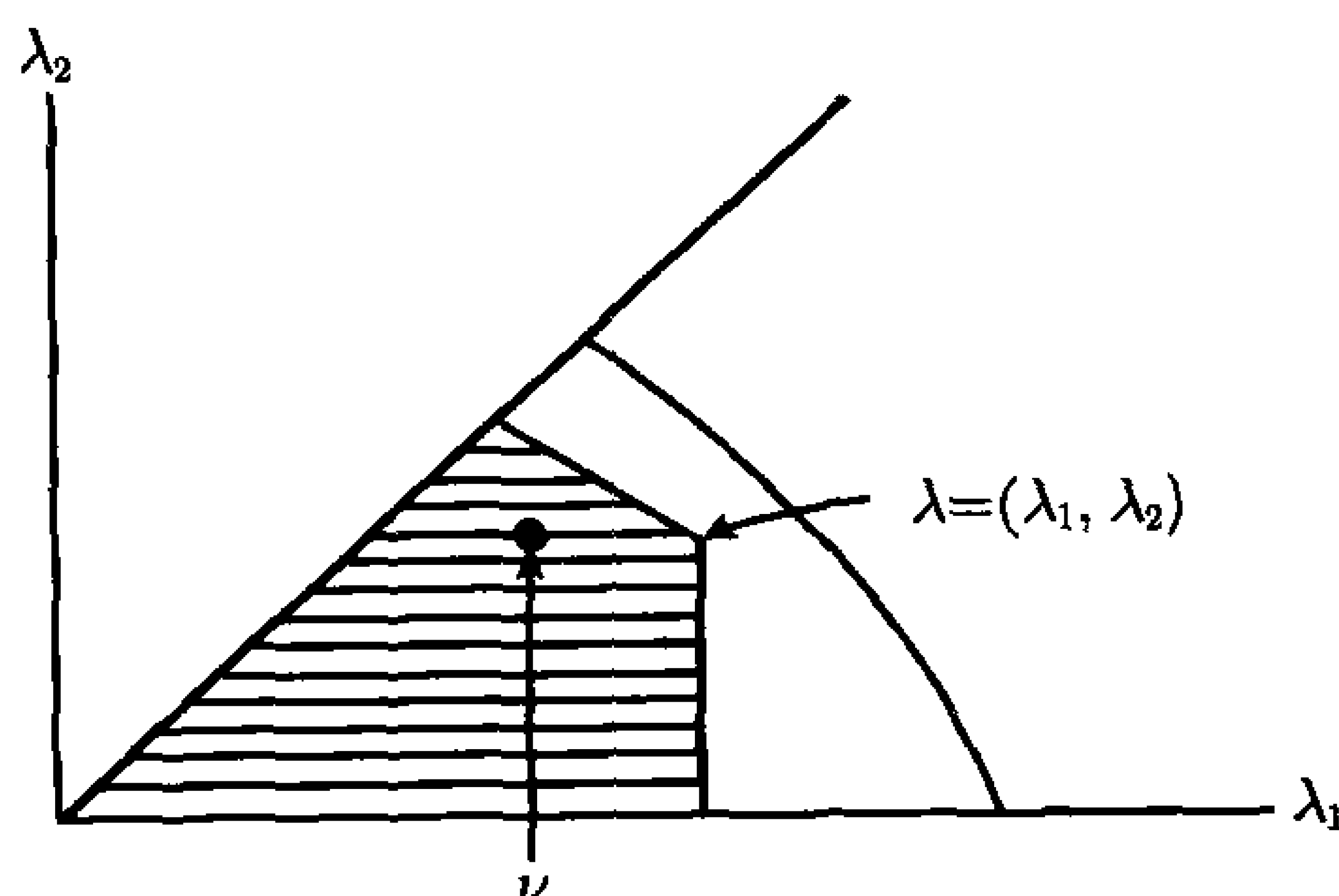


图 8.3 以优势单调的区域

定理 8.10.2 如果区域 $A \subset R^m_<$ 是闭的, 凸的且以优势单调的, 则 A 是某容许检验的接受区域.

定理 8.10.1 和定理 8.10.2 是等价的, 先证定理 8.10.2 会容易些. 利用某一个凸集的极值点 (引理 8.10.11) 可证这两个定理的等价性.

因为 $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ 的分布可以写成指数形式, 所以能够应用定理 5.6.5(Stein 定理). 设 $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{X}' + \boldsymbol{Y}\boldsymbol{Y}' + \boldsymbol{Z}\boldsymbol{Z}' = (u_{ij})$ 以及 $\boldsymbol{\Sigma}^{-1} = (\sigma_{ij})$. 对于一般的矩阵 $\boldsymbol{C} = (\boldsymbol{c}_1, \cdots, \boldsymbol{c}_k)$, 设 $\mathrm{vec}(\boldsymbol{C}) = (\boldsymbol{c}_1', \cdots, \boldsymbol{c}_k')'$. $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ 的密度可以写为

$$\begin{aligned} f(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) &= K(\boldsymbol{\Xi}, \boldsymbol{H}, \boldsymbol{\Sigma}) \exp\{\mathrm{tr}\boldsymbol{\Xi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} + \mathrm{tr}\boldsymbol{H}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} - \frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}\} \\ &= K(\boldsymbol{\Xi}, \boldsymbol{H}, \boldsymbol{\Sigma}) \exp\{\boldsymbol{\omega}_{(1)}'\boldsymbol{y}_{(1)} + \boldsymbol{\omega}_{(2)}'\boldsymbol{y}_{(2)} + \boldsymbol{\omega}_{(3)}'\boldsymbol{y}_{(3)}\}, \end{aligned} \tag{9}$$

其中 $K(\boldsymbol{\Xi}, \boldsymbol{H}, \boldsymbol{\Sigma})$ 是常量,

$$\begin{aligned} &\boldsymbol{\omega}_{(1)} = \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}), \quad \boldsymbol{\omega}_{(2)} = \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\boldsymbol{H}), \\ &\boldsymbol{\omega}_{(3)} = -\frac{1}{2}(\sigma^{11}, 2\sigma^{12}, \cdots, 2\sigma^{1p}, \sigma^{22}, \cdots, \sigma^{pp})', \\ &\boldsymbol{y}_{(1)} = \mathrm{vec}(\boldsymbol{X}), \quad \boldsymbol{y}_{(2)} = \mathrm{vec}(\boldsymbol{Y}), \\ &\boldsymbol{y}_{(3)} = (u_{11}, u_{12}, \cdots, u_{1p}, u_{22}, \cdots, u_{pp})'. \end{aligned} \tag{10}$$

如果用 g 表示映射 $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \to \boldsymbol{y} = (\boldsymbol{y}_{(1)}', \boldsymbol{y}_{(2)}', \boldsymbol{y}_{(3)}')'$, 那么 $\boldsymbol{y} = g(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ 而且在$\boldsymbol{y}$空间中集合 A 的测度为 $m(A) = \mu(g^{-1}(A))$, 其中 μ 为 $R^{p(m+r+n)}$ 上的一般的勒贝格测度. 由于 $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U})$ 是充分统计量, 所以 $\boldsymbol{y} = (\boldsymbol{y}_{(1)}', \boldsymbol{y}_{(2)}', \boldsymbol{y}_{(3)}')'$ 也是充分统计量. 因为若一个检验关于一个基于充分统计量的检验类是容许的, 则它关于

所有的检验类也是容许的, 所以我们只考虑基于充分统计量的检验. 从而这些检验的接受区域都是 $\boldsymbol{y}$ 空间中的子集. 由于 (9) 式右端给出的 $\boldsymbol{y}$ 的密度具有指数分布族的形式, 因此我们可以利用 Stein 定理. 另外, 因为变换 $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U}) \to \boldsymbol{y}$ 是线性的, 我们还要证明 $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U})$ 的接受区域是凸的. 一个不变检验的接受区域是基于 $\boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V})) = (\lambda_1,\cdots,\lambda_m)'$ 给出的, 所以为了证明这些检验的容许性, 我们需要验证 A 的逆象, 即 $\tilde{A} = \{\boldsymbol{V}|\boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V})) \in A\}$, 满足 Stein 定理中的条件, 也就是说它是一个凸集.

设 $\boldsymbol{V}_i = (\boldsymbol{X}_i,\boldsymbol{Y}_i,\boldsymbol{U}_i) \in \tilde{A}, i = 1,2$, 即 $\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_i)] \in A$. 由 A 的凸性可得当 $0 \leqslant p = 1-q \leqslant 1$ 时, $p\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_1)] + q\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_2)] \in A$. 为了证明 $p\boldsymbol{V}_1 + q\boldsymbol{V}_2 \in \tilde{A}$, 即 $\boldsymbol{\lambda}[\boldsymbol{M}(p\boldsymbol{V}_1 + q\boldsymbol{V}_2)] \in A$, 我们利用 A 的优势单调性和下面的定理.

定理 8.10.3

$$\boldsymbol{\lambda}[\boldsymbol{M}(p\boldsymbol{V}_1 + q\boldsymbol{V}_2)] \succ_w p\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_1)] + q\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_2)]. \tag{11}$$

定理 8.10.3 (图 8.4) 由下式可证

$$\begin{aligned}\boldsymbol{\lambda}[\boldsymbol{M}(p\boldsymbol{V}_1 + q\boldsymbol{V}_2)] &\succ_w \boldsymbol{\lambda}[p\boldsymbol{M}(\boldsymbol{V}_1) + q\boldsymbol{M}(\boldsymbol{V}_2)] \\ &\succ_w p\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_1)] + q\boldsymbol{\lambda}[\boldsymbol{M}(\boldsymbol{V}_2)].\end{aligned} \tag{12}$$

(12) 式中的第二个优势结论是下面引理的特殊情况.

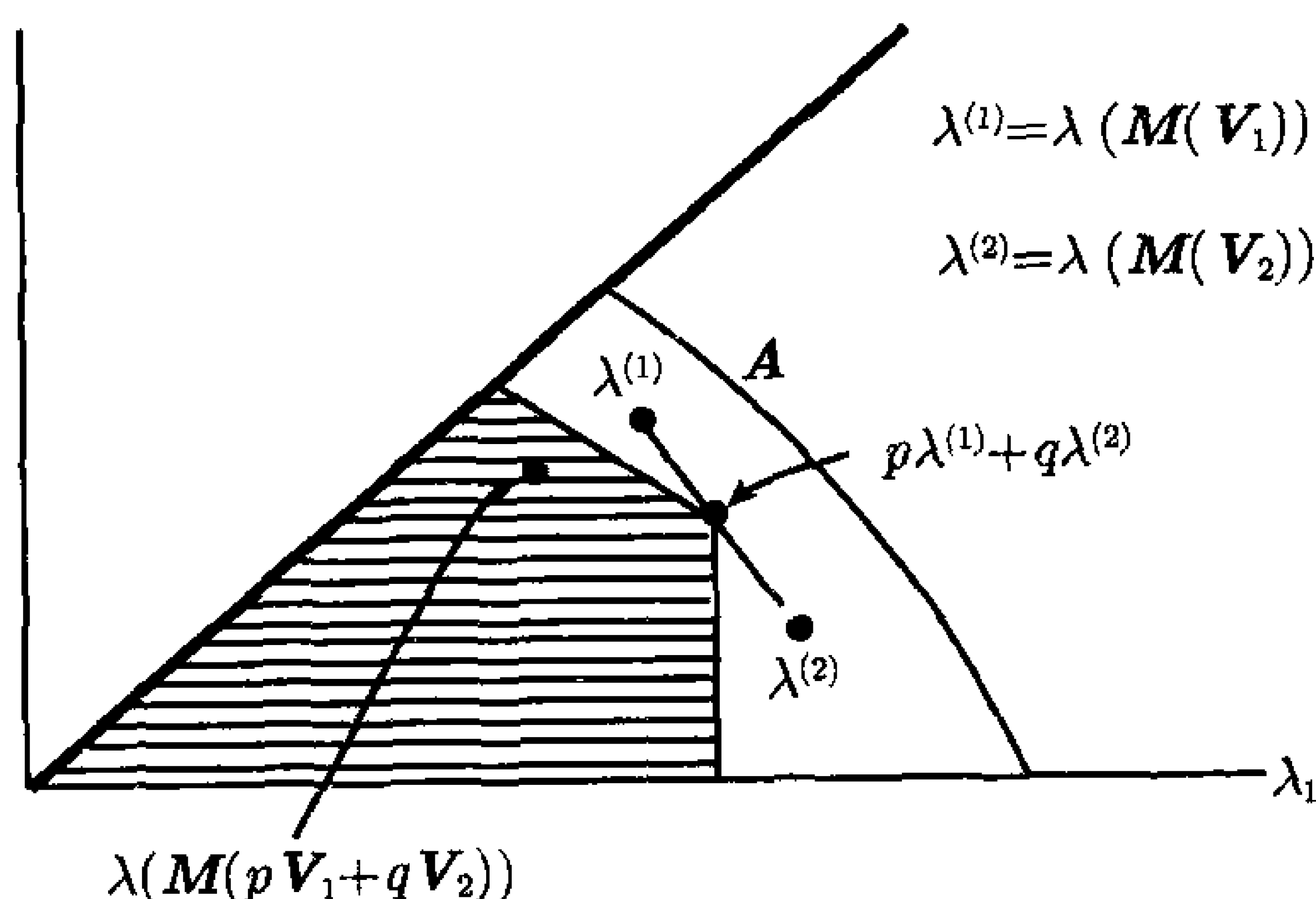


图 8.4　定理 8.10.3

引理 8.10.1　*若$\boldsymbol{A}$和$\boldsymbol{B}$对称, 则有*

$$\boldsymbol{\lambda}(\boldsymbol{A}+\boldsymbol{B}) \succ_w \boldsymbol{\lambda}(\boldsymbol{A}) + \boldsymbol{\lambda}(\boldsymbol{B}). \tag{13}$$

证明　由附录中的推论 A.4.2 可得

$$\begin{aligned}\sum_{i=1}^{k} \lambda_i(\boldsymbol{A}+\boldsymbol{B}) &= \max_{\boldsymbol{R}'\boldsymbol{R}=\boldsymbol{I}_k} \operatorname{tr}\boldsymbol{R}'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{R} \\ &\leqslant \max_{\boldsymbol{R}'\boldsymbol{R}=\boldsymbol{I}_k} \operatorname{tr}\boldsymbol{R}'\boldsymbol{A}\boldsymbol{R} + \max_{\boldsymbol{R}'\boldsymbol{R}=\boldsymbol{I}_k} \operatorname{tr}\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}\end{aligned} \tag{14}$$

$$=\sum_{i=1}^{k}\lambda_i(\boldsymbol{A})+\sum_{i=1}^{k}\lambda_i(\boldsymbol{B})$$
$$=\sum_{i=1}^{k}\{\lambda_i(\boldsymbol{A})+\lambda_i(\boldsymbol{B})\},\qquad k=1,\cdots,p.$$ ■

令 $\boldsymbol{A}>\boldsymbol{B}$ 表示 $\boldsymbol{A}-\boldsymbol{B}$ 是正定的, $\boldsymbol{A}\geqslant\boldsymbol{B}$ 表示 $\boldsymbol{A}-\boldsymbol{B}$ 是半正定的.

由下面这些引理可得 (12) 式的第一个优势结论.

引理 8.10.2

$$p\boldsymbol{U}_1+q\boldsymbol{U}_2-(p\boldsymbol{Y}_1+q\boldsymbol{Y}_2)(p\boldsymbol{Y}_1+q\boldsymbol{Y}_2)' \tag{15}$$
$$\geqslant p(\boldsymbol{U}_1-\boldsymbol{Y}_1\boldsymbol{Y}_1')+q(\boldsymbol{U}_2-\boldsymbol{Y}_2\boldsymbol{Y}_2').$$

证明 左端减去右端可得

$$p\boldsymbol{Y}_1\boldsymbol{Y}_1'+q\boldsymbol{Y}_2\boldsymbol{Y}_2'-p^2\boldsymbol{Y}_1\boldsymbol{Y}_1'-q^2\boldsymbol{Y}_2\boldsymbol{Y}_2'-pq(\boldsymbol{Y}_1\boldsymbol{Y}_2'+\boldsymbol{Y}_2\boldsymbol{Y}_1') \tag{16}$$
$$=p(1-p)\boldsymbol{Y}_1\boldsymbol{Y}_1'+q(1-q)\boldsymbol{Y}_2\boldsymbol{Y}_2'-pq(\boldsymbol{Y}_1\boldsymbol{Y}_2'+\boldsymbol{Y}_2\boldsymbol{Y}_1')$$
$$=pq(\boldsymbol{Y}_1-\boldsymbol{Y}_2)(\boldsymbol{Y}_1-\boldsymbol{Y}_2)'\geqslant\boldsymbol{0}.$$ ■

引理 8.10.3 若 $\boldsymbol{A}\geqslant\boldsymbol{B}>\boldsymbol{0}$, 则 $\boldsymbol{A}^{-1}\leqslant\boldsymbol{B}^{-1}$.

证明 见习题 8.31. ■

引理 8.10.4 若 $\boldsymbol{A}>\boldsymbol{0}$, 则 $f(\boldsymbol{x},\boldsymbol{A})=\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}$ 在 $(\boldsymbol{x},\boldsymbol{A})$ 中是凸的.

证明 见习题 5.17. ■

引理 8.10.5 若 $\boldsymbol{A}_1>\boldsymbol{0}$, $\boldsymbol{A}_2>\boldsymbol{0}$, 则

$$(p\boldsymbol{B}_1+q\boldsymbol{B}_2)'(p\boldsymbol{A}_1+q\boldsymbol{A}_2)^{-1}(p\boldsymbol{B}_1+q\boldsymbol{B}_2)\leqslant p\boldsymbol{B}_1'\boldsymbol{A}_1^{-1}\boldsymbol{B}_1+q\boldsymbol{B}_2'\boldsymbol{A}_2^{-1}\boldsymbol{B}_2. \tag{17}$$

证明 由引理 8.10.4 可知, 对所有的$\boldsymbol{y}$有

$$p\boldsymbol{y}'\boldsymbol{B}_1'\boldsymbol{A}_1^{-1}\boldsymbol{B}_1\boldsymbol{y}+q\boldsymbol{y}'\boldsymbol{B}_2'\boldsymbol{A}_2^{-1}\boldsymbol{B}_2\boldsymbol{y} \tag{18}$$
$$-\boldsymbol{y}'(p\boldsymbol{B}_1+q\boldsymbol{B}_2)'(p\boldsymbol{A}_1+q\boldsymbol{A}_2)^{-1}(p\boldsymbol{B}_1+q\boldsymbol{B}_2)\boldsymbol{y}$$
$$=p(\boldsymbol{B}_1\boldsymbol{y})'\boldsymbol{A}_1^{-1}(\boldsymbol{B}_1\boldsymbol{y})+q(\boldsymbol{B}_2\boldsymbol{y})'\boldsymbol{A}_2^{-1}(\boldsymbol{B}_2\boldsymbol{y})$$
$$-(p\boldsymbol{B}_1\boldsymbol{y}+q\boldsymbol{B}_2\boldsymbol{y})'(p\boldsymbol{A}_1+q\boldsymbol{A}_2)^{-1}(p\boldsymbol{B}_1\boldsymbol{y}+q\boldsymbol{B}_2\boldsymbol{y})$$
$$\geqslant 0.$$

因此关于 $\boldsymbol{y}$ 的二次型的矩阵是半正定的. ■

(17) 式中的凸性有时也称为矩阵的凸性. [见 Marshall and Olkin (1979).]

引理 8.10.6

$$\boldsymbol{M}(p\boldsymbol{V}_1+q\boldsymbol{V}_2)\leqslant p\boldsymbol{M}(\boldsymbol{V}_1)+q\boldsymbol{M}(\boldsymbol{V}_2), \tag{19}$$

其中 $\boldsymbol{V}_1=(\boldsymbol{X}_1,\boldsymbol{Y}_1,\boldsymbol{U}_1),\boldsymbol{V}_2=(\boldsymbol{X}_2,\boldsymbol{Y}_2,\boldsymbol{U}_2),\boldsymbol{U}_1-\boldsymbol{Y}_1\boldsymbol{Y}_1'>\boldsymbol{0},\boldsymbol{U}_2-\boldsymbol{Y}_2\boldsymbol{Y}_2'>\boldsymbol{0},0\leqslant p=1-q\leqslant 1$.

证明 由引理 8.10.2 和引理 8.10.3 可知

$$[p\boldsymbol{U}_1+q\boldsymbol{U}_2-(p\boldsymbol{Y}_1+q\boldsymbol{Y}_2)(p\boldsymbol{Y}_1+q\boldsymbol{Y}_2)']^{-1} \leqslant [p(\boldsymbol{U}_1-\boldsymbol{Y}_1\boldsymbol{Y}_1')+q(\boldsymbol{U}_2-\boldsymbol{Y}_2\boldsymbol{Y}_2')]^{-1}. \tag{20}$$

从而

$$\boldsymbol{M}(p\boldsymbol{V}_1+q\boldsymbol{V}_2) \leqslant (p\boldsymbol{X}_1+q\boldsymbol{X}_2)'[p(\boldsymbol{U}_1-\boldsymbol{Y}_1\boldsymbol{Y}_1')+q(\boldsymbol{U}_2-\boldsymbol{Y}_2\boldsymbol{Y}_2')]^{-1}(p\boldsymbol{X}_1+q\boldsymbol{X}_2). \tag{21}$$

由引理 8.10.5 可知, (21) 式的右端小于等于

$$p\boldsymbol{X}_1'(\boldsymbol{U}_1-\boldsymbol{Y}_1\boldsymbol{Y}_1')^{-1}\boldsymbol{X}_1+q\boldsymbol{X}_2'(\boldsymbol{U}_2-\boldsymbol{Y}_2\boldsymbol{Y}_2')^{-1}\boldsymbol{X}_2=p\boldsymbol{M}(\boldsymbol{V}_1)+q\boldsymbol{M}(\boldsymbol{V}_2). \ \blacksquare \tag{22}$$

引理 8.10.7　*若 $\boldsymbol{A}\leqslant\boldsymbol{B}$, 则 $\boldsymbol{\lambda}(\boldsymbol{A})\prec_w\boldsymbol{\lambda}(\boldsymbol{B})$.*

证明　由附录中的推论 A.4.2 可得,

$$\sum_{i=1}^k\lambda_i(\boldsymbol{A})=\max_{\boldsymbol{R}'\boldsymbol{R}=\boldsymbol{I}_k}\mathrm{tr}\boldsymbol{R}'\boldsymbol{A}\boldsymbol{R}\leqslant\max_{\boldsymbol{R}'\boldsymbol{R}=\boldsymbol{I}_k}\mathrm{tr}\boldsymbol{R}'\boldsymbol{B}\boldsymbol{R}=\sum_{i=1}^k\lambda_i(\boldsymbol{B}),\quad k=1,\cdots,p. \ \blacksquare \tag{23}$$

由引理 8.10.7 可得 (12) 式中的第一个优势结论, 从而可得定理 8.10.3 和 $\tilde{A}$ 的凸性. 因此接受区域满足 Stein 定理中的条件 (i).

引理 8.10.8　*定理 8.10.1 或定理 8.10.2 中的接受区域 A 满足 Stein 定理中的条件 (ii).*

证明　设 ω 对应于 $(\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Theta})$, 则

$$\begin{aligned}\boldsymbol{\omega}'\boldsymbol{y}&=\boldsymbol{\omega}_{(1)}'\boldsymbol{y}_{(1)}+\boldsymbol{\omega}_{(2)}'\boldsymbol{y}_{(2)}+\boldsymbol{\omega}_{(3)}'\boldsymbol{y}_{(3)}\\&=\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{X}+\mathrm{tr}\boldsymbol{\Psi}'\boldsymbol{Y}-\frac{1}{2}\mathrm{tr}\boldsymbol{\Theta}\boldsymbol{U},\end{aligned} \tag{24}$$

其中$\boldsymbol{\Theta}$是对称的. 假设 $\{\boldsymbol{y}|\boldsymbol{\omega}'\boldsymbol{y}>c\}$ 与 $\tilde{A}=\{\boldsymbol{V}|\boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V}))\in A\}$ 不相交, 我们要证明这时的$\boldsymbol{\Theta}$是半正定的. 如果$\boldsymbol{\Theta}$不是半正定的, 则

$$\boldsymbol{\Theta}=\boldsymbol{D}\begin{bmatrix}\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}\\\boldsymbol{0}&-\boldsymbol{I}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}\end{bmatrix}\boldsymbol{D}', \tag{25}$$

其中$\boldsymbol{D}$是非奇异的, $-\boldsymbol{I}$ 是非空的. 令 $\boldsymbol{X}=(1/\gamma)\boldsymbol{X}_0, \boldsymbol{Y}=(1/\gamma)\boldsymbol{Y}_0$,

$$\boldsymbol{U}=(\boldsymbol{D}')^{-1}\begin{pmatrix}\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}\\\boldsymbol{0}&\gamma\boldsymbol{I}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{I}\end{pmatrix}\boldsymbol{D}^{-1}, \tag{26}$$

以及 $\boldsymbol{V}=(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U})$, 其中 $\boldsymbol{X}_0,\boldsymbol{Y}_0$ 是固定的矩阵, γ 是一个正常数. 则对充分大的 γ, 有下式成立

$$\boldsymbol{\omega}'\boldsymbol{y}=\frac{1}{\gamma}\mathrm{tr}\boldsymbol{\Phi}'\boldsymbol{X}_0+\frac{1}{\gamma}\mathrm{tr}\boldsymbol{\Psi}'\boldsymbol{Y}_0+\frac{1}{2}\mathrm{tr}\begin{pmatrix}-\boldsymbol{I}&\boldsymbol{0}&\boldsymbol{0}\\\boldsymbol{0}&\gamma\boldsymbol{I}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}&\boldsymbol{0}\end{pmatrix}>c. \tag{27}$$

另一方面, 当 $\gamma \to \infty$ 时

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{M}(\boldsymbol{V})) &= \boldsymbol{\lambda}\{\boldsymbol{X}'(\boldsymbol{U}-\boldsymbol{Y}\boldsymbol{Y}')^{-1}\boldsymbol{X}\} \\ &= \frac{1}{\gamma^2}\boldsymbol{\lambda}\left\{\boldsymbol{X}_0'\left[(\boldsymbol{D}')^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \gamma\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}\boldsymbol{D}^{-1}-\frac{1}{\gamma^2}\boldsymbol{Y}_0\boldsymbol{Y}_0'\right]^{-1}\boldsymbol{X}_0\right\} \\ &\to \boldsymbol{0}.\end{aligned} \tag{28}$$

因此, 对充分大的 γ 有 $\boldsymbol{V} \in \tilde{A}$. 这是矛盾的. 所以$\boldsymbol{\Theta}$为半正定矩阵.

令 ω_1 对应于 $(\boldsymbol{\Phi}_1, \boldsymbol{0}, \boldsymbol{I})$, 其中 $\boldsymbol{\Phi}_1 \neq \boldsymbol{0}$. 那么 $\boldsymbol{I}+\lambda\boldsymbol{\Theta}$ 是正定矩阵, 且对充分大的 λ 有 $\boldsymbol{\Phi}_1+\lambda\boldsymbol{\Phi} \neq \boldsymbol{0}$. 因此对充分大的 λ 有 $\omega_1+\lambda\omega \in \Omega-\Omega_0$.

前面的证明由 Charles Stein 给出. ■

现在由定理 5.6.5、定理 8.10.3 和引理 8.10.8 可得定理 8.10.2.

为了由定理 8.10.2 得到定理 8.10.1, 我们需要下面的引理.

引理 8.10.9 $A \subset R_{\leq}^m$ **是凸的且以优势单调的当且仅当** A **是单调的且** A^* **是凸的.**

证明 必要性. 若 A 以优势单调, 则显然是单调的. A^* 是凸的 (见习题 8.35).

充分性. 对于 $\lambda \subset R_{\leq}^m$, 设

$$\begin{aligned}C(\boldsymbol{\lambda}) &= \{\boldsymbol{x}|\boldsymbol{x} \in R_+^m, \boldsymbol{x} \succ_w \boldsymbol{\lambda}\}, \\ D(\boldsymbol{\lambda}) &= \{\boldsymbol{x}|\boldsymbol{x} \in R_{\leq}^m, \boldsymbol{x} \succ_w \boldsymbol{\lambda}\}.\end{aligned} \tag{29}$$

由引理 8.10.10, 引理 8.10.11 和它的推论可得, A 的单调性和 A^* 的凸性隐含了 $C(\boldsymbol{\lambda}) \subset A^*$. 所以 $D(\boldsymbol{\lambda}) = C(\boldsymbol{\lambda}) \cap R_{\leq}^m \subset A^* \cap R_{\leq}^m = A$. 现在我们假设 $\boldsymbol{\nu} \in R_{\leq}^m$ 和 $\boldsymbol{\nu} \prec_w \boldsymbol{\lambda}$. 这时 $\boldsymbol{\nu} \in D(\boldsymbol{\lambda}) \subset A$. 这说明 A 是以优势单调的. 另外, 若 A^* 是凸的, 则 $A = R_{\leq}^m \cap A^*$ 也是凸的.(见图 8.5.) ■

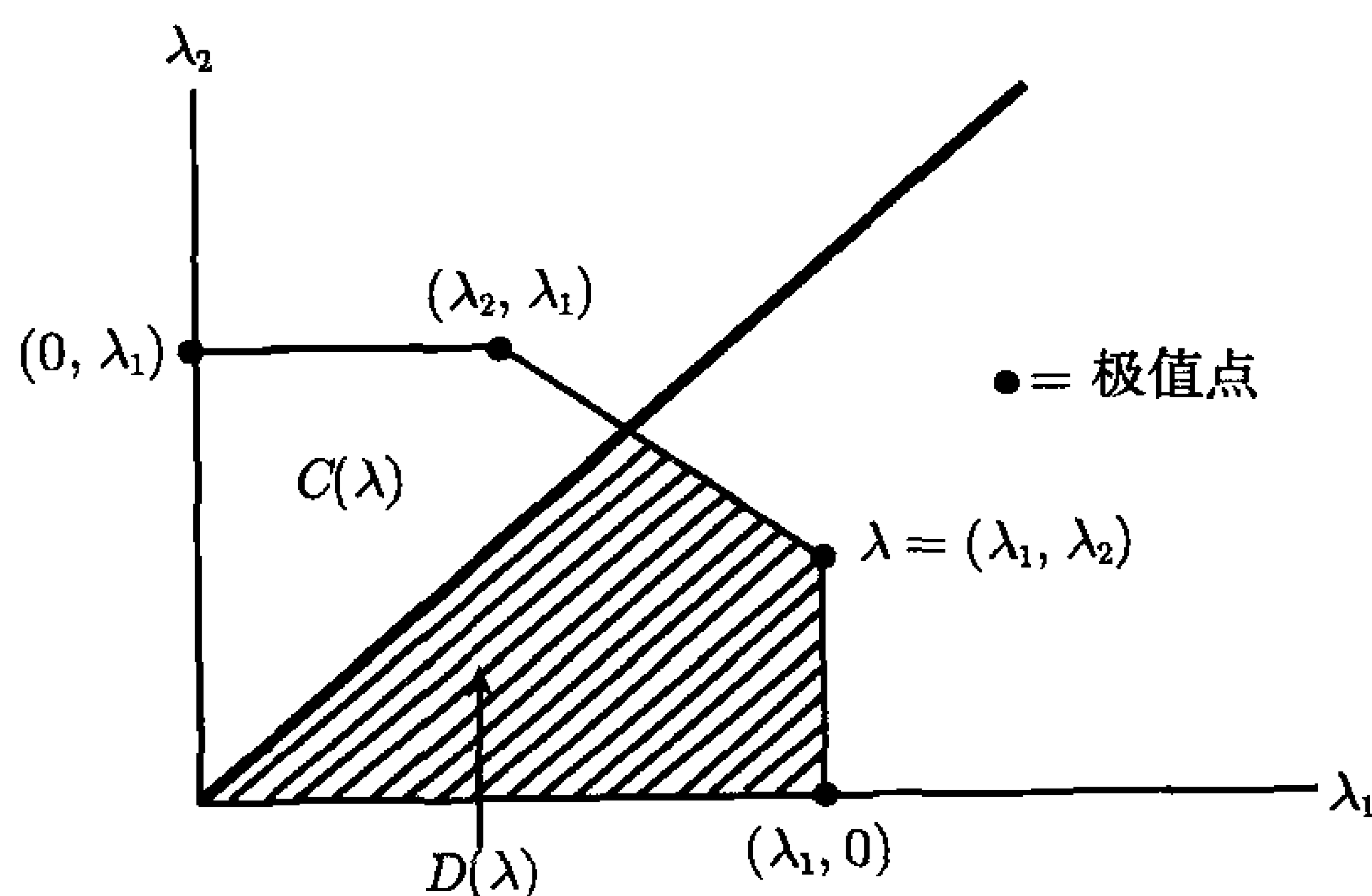


图 8.5

引理 8.10.10 设 C 是紧的而且是凸的, D 是凸的. 若 C 的极值点包含在 D 内, 则 $C \subset D$.

证明 显然. ■

引理 8.10.11 $C(\boldsymbol{\lambda})$ 的每一个极值点都形如

$$(\delta_{\pi(1)}\lambda_{\pi(1)}, \cdots, \delta_{\pi(m)}\lambda_{\pi(m)}), \tag{30}$$

其中 π 是 $(1, \cdots, m)$ 的置换, 且存在某个 k 使得 $\delta_1 = \cdots = \delta_k = 1, \delta_{k+1} = \cdots = \delta_m = 0$.

证明 $C(\boldsymbol{\lambda})$ 是凸的.(见习题 8.34.) 注意到 $C(\boldsymbol{\lambda})$ 是对称置换, 即, 若 $(x_1, \cdots, x_m)' \in C(\boldsymbol{\lambda})$, 则对任意的置换 π 有 $(x_{\pi(1)}, \cdots, x_{\pi(m)})' \in C(\boldsymbol{\lambda})$. 因此对任意一个置换 π, $\pi(C(\boldsymbol{\lambda})) = \{(x_{\pi(1)}, \cdots, x_{\pi(m)})' | \boldsymbol{x} \in C(\boldsymbol{\lambda})\}$ 与 $C(\boldsymbol{\lambda})$ 重合. 由此可知, 若 $(x_1, \cdots, x_m)'$ 是 $C(\boldsymbol{\lambda})$ 的极值点, 则 $(x_{\pi(1)}, \cdots, x_{\pi(m)})'$ 也是一个极值点. 特别地 $(x_{[1]}, \cdots, x_{[m]}) \in R^m_<$ 是 $C(\boldsymbol{\lambda})$ 的一个极值点. 反之, 若 $(x_1, \cdots, x_m) \in R^m_<$ 是 $C(\boldsymbol{\lambda})$ 的一个极值点, 则 $(x_{\pi(1)}, \cdots, x_{\pi(m)})'$ 也是一个极值点.

我们看到, 一旦列举出 $C(\boldsymbol{\lambda})$ 在 $R^m_<$ 中的极值点, 其余的极值点可以通过置换获得.

设 $\boldsymbol{x} \in R^m_<$. 一个极点, 同时也是 m 个超平面的交点, 它必满足下面 $2m$ 个方程中的至少 m 个方程:

$$\begin{aligned} &E_1 : x_1 = 0, \quad &&F_1 : x_1 = \lambda_1, \\ &E_2 : x_2 = 0, \quad &&F_2 : x_1 + x_2 = \lambda_1 + \lambda_2, \\ &\quad\vdots &&\quad\vdots \\ &E_m : x_m = 0, \quad &&F_m : x_1 + \cdots + x_m = \lambda_1 + \cdots + \lambda_m. \end{aligned} \tag{31}$$

设 k 是使 E_k 成立的第一个下标. 则由 $\boldsymbol{x} \in R^m_<$ 可得 $0 = x_k \geqslant x_{k+1} \geqslant \cdots \geqslant x_m \geqslant 0$. 因此 $E_k, \cdots, E_m$ 成立. 剩下的 $k-1 = m-(m-k+1)$ 个或更多个方程在 F 中. 我们把它们编号为 $F_{i_1}, \cdots, F_{i_l}$, 其中 $i_1 < \cdots < i_l, l \geqslant k-1$. 由 $i_1 < \cdots < i_l$ 可得 $i_1 \geqslant l$, 等号当且仅当 $i_1 = 1, \cdots, i_l = l$ 时成立. 此时 $F_1. \cdots, F_{k-1}\,(l \geqslant k-1)$ 成立. 下面设 $i_l > l$. 因为 $x_k = \cdots = x_m = 0$, 所以

$$F_{i_l} : x_1 + \cdots + x_{k-1} = \lambda_1 + \cdots + \lambda_{k-1} + \cdots + \lambda_{i_l}. \tag{32}$$

但是 $x_1 + \cdots + x_{k-1} \leqslant \lambda_1 + \cdots + \lambda_{k-1}$, 我们有 $\lambda_k + \cdots + \lambda_{i_l} = 0$. 因此 $0 = \lambda_k + \cdots + \lambda_{i_l} \geqslant \lambda_k \geqslant \cdots \geqslant \lambda_m \geqslant 0$. 此时 $F_{k-1}, \cdots, F_m$ 变为同一个方程 $x_1 + \cdots + x_{k-1} = \lambda_1 + \cdots + \lambda_{k-1}$. 从而 $\boldsymbol{x}$ 满足另外 $k-2$ 个方程, 且这 $k-2$ 个方程必定是 $F_1. \cdots, F_{k-2}$. 这就证明了无论是哪种情况, $E_k, \cdots, E_m, F_1. \cdots, F_{k-1}$ 都成立, 这就给出了 $R^m_< \cap C(\boldsymbol{\lambda})$ 中的点 $\beta = (\lambda_1, \cdots, \lambda_{k-1}, 0, \cdots, 0)$. 因此 β 是一个极值点. ■

推论 8.10.1 $C(\boldsymbol{\lambda}) \in A^*$.

证明 若 A 是单调的, 那么如果 $\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_m)'\in A^*,\boldsymbol{\nu}=(\nu_1,\cdots,\nu_m)',\nu_i\leqslant\lambda_i,i=1,\cdots,m$, 则 A^* 是单调的, 所以 $\boldsymbol{\nu}\in A^*$.(见习题 8.35.) 由置换的对称性和 A^* 的单调性可知, 由 (30) 给出的 $C(\boldsymbol{\lambda})$ 的极值点在 A^* 中. 因此, 由引理 8.10.10 可知, $C(\boldsymbol{\lambda})\in A^*$. ■

定理 8.10.1 的证明 由定理 8.10.2 和引理 8.10.9 直接可得. ■

应用 Schur 凸函数, 可得定理 8.10.2 的几个推论.

推论 8.10.2 *设 g 在 $[0,1)$ 上是连续的、非降的且是凸的. 设*

$$f(\boldsymbol{\lambda})=f(\lambda_1,\cdots,\lambda_m)=\sum_{i=1}^{m}g(\lambda_i). \tag{33}$$

则以 $A=\{\boldsymbol{\lambda}|f(\boldsymbol{\lambda})\leqslant c\}$ 为接受区域的检验是容许的.

证明 凸函数的和 f 还是凸的, 因此 A 是凸的. 因为 f 是连续的, 所以 A 是闭的. 我们想证明, 若 $f(\boldsymbol{x})\leqslant c$ 且 $\boldsymbol{y}\prec_w\boldsymbol{x}\,(\boldsymbol{x},\boldsymbol{y}\in R^m_{\leqslant})$, 则 $f(\boldsymbol{y})\leqslant c$. 设 $\tilde{x}_k=\sum_{i=1}^k x_i,\tilde{y}_k=\sum_{i=1}^k y_i$. 则 $\boldsymbol{y}\prec_w\boldsymbol{x}$ 当且仅当 $\tilde{x}_k\geqslant\tilde{y}_k,k=1,\cdots,m$. 令 $f(\boldsymbol{x})=h(\tilde{x}_1,\cdots,\tilde{x}_m)=g(\tilde{x}_1)+\sum_{i=2}^m g(\tilde{x}_i-\tilde{x}_{i-1})$, 我们只需要证明 $h(\tilde{x}_1,\cdots,\tilde{x}_m)$ 关于每一个 $\tilde{x}_i$ 递增. 当 $i\leqslant m-1$ 时, 由 g 的凸性可得,

$$\begin{aligned}&h(\tilde{x}_1,\cdots,\tilde{x}_i+\varepsilon,\cdots,\tilde{x}_m)-h(\tilde{x}_1,\cdots,\tilde{x}_i,\cdots,\tilde{x}_m)\\&=g(x_1+\varepsilon)-g(x_i)-\{g(x_{i+1})-g(x_{i+1}-\varepsilon)\}\geqslant 0.\end{aligned} \tag{34}$$

当 $i=m$ 时, 由 g 的单调性可知,

$$h(\tilde{x}_1,\cdots,\tilde{x}_m+\varepsilon)-h(\tilde{x}_1,\cdots,\tilde{x}_m)=g(x_m+\varepsilon)-g(x_m)\geqslant 0. \quad ■ \tag{35}$$

分别令 $g(\lambda)=-\ln(1-\lambda),g(\lambda)=\lambda/(1-\lambda),g(\lambda)=\lambda$, 易证 Wilks 似然比检验、Lawley-Hotelling 迹检验和 Bartlett-Nanda-Pillai 检验是容许的. 由定理 8.10.1 或定理 8.10.2 直接可得 Roy 最大根检验 $A:\lambda_1\leqslant c$ 的容许性. 然而最小根检验 $\lambda_t\leqslant c$, 其中 $t=\min(m,p)$, 不满足凸性条件. 下面的定理证明了该检验实际上是不容许的.

定理 8.10.4 *不变检验是容许检验的必要条件为: $\sqrt{\lambda_1},\cdots,\sqrt{\lambda_t}$ 空间中的扩展区域是凸的而且是单调的.*

我们概括地叙述一下该定理的证明 [来自于 Schwartz (1967)]. 令 $\sqrt{\lambda_i}=d_i,i=1,\cdots,t$, 并设 $d_1,\cdots,d_t$ 的密度为 $f(\boldsymbol{d}|\boldsymbol{\nu})$, 其中 $\boldsymbol{\nu}=(\nu_1,\cdots,\nu_t)'$ 的定义在 8.6.5 节给出, $f(\boldsymbol{d}|\boldsymbol{\nu})$ 由第 13 章给出的.

比值 $f(\boldsymbol{d}|\boldsymbol{\nu})/f(\boldsymbol{d}|\boldsymbol{0})$ 可以对称地推广到单位立方体 $(0\leqslant d_i\leqslant 1,i=1,\cdots,t)$. 推广的比值是凸函数且关于每个 d_i 是严格递增的. 一个恰当的贝叶斯过程的接受区域为

$$\int\frac{f(\boldsymbol{d}|\boldsymbol{\nu})}{f(\boldsymbol{d}|\boldsymbol{0})}d\Pi(\boldsymbol{\nu})\leqslant c, \tag{36}$$

其中 $\Pi(\boldsymbol{\nu})$ 是 $\boldsymbol{\nu}$ 空间上的有限测度. 那么满足 (36) 的$\boldsymbol{d}$的集合的对称推广是凸的且单调的 [如 Birnbaum(1955) 所证]. 贝叶斯过程的集合的闭包 (在弱 * 拓扑下) 形成一个本质完全类 [Wald(1950)]. 此时, 凸单调接受区域的极限是凸的且单调的. 这里关于容许性的讨论是由 Anderson and Takemura (1982) 阐述的.

8.10.2 检验的无偏性和功效函数的单调性

如果检验 T 的功效在原假设成立时达到极小值, 则该检验是**无偏的**. 如果在参数空间中定义了自然参数化法和距离的概念, 而且功效函数随着备择假设和原假设之间的距离的增加而增加, 则称该功效函数为**单调的**. 单调性意味着无偏性. 本节我们将证明一般线性假设的许多不变检验的功效函数在参数 (即根) 不变时是单调的. 这些都可以作为距离的测度.

我们首先考虑检验原假设 $\mu = 0$ 和备择假设 $\mu \neq 0$ 的接受区间 $(-a, a)$, 其观测值来自 $N(\mu, \sigma^2)$. 在图 8.6 中, 对于 μ 的三个值, 我们用阴影区域表示接受概率. 显然, 接受概率随着 μ 远离零单调递减 (或等价地, 功效单调递增). 实际上这个性质只依赖于密度函数的单峰性和对称性.

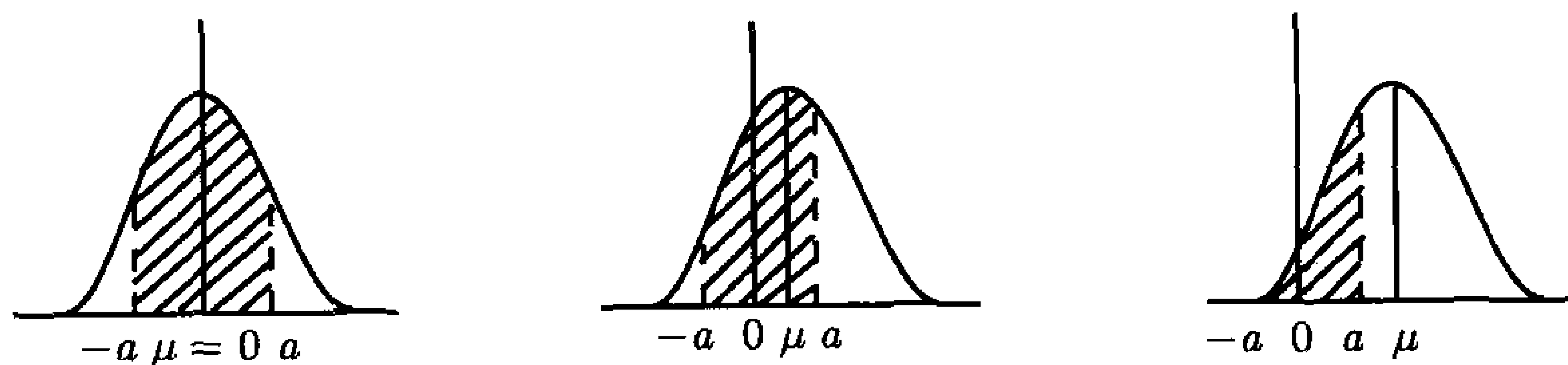


图 8.6 三个接受概率

对高维的情况, 我们把区间推广为对称的凸集, 并要求密度函数具有对称性和单峰性, 即常密度的每个周线线环绕一个凸集. 由图 8.7 可见, 此时的接受概率单调递减. 下面的定理由 Anderson(1955b) 给出.

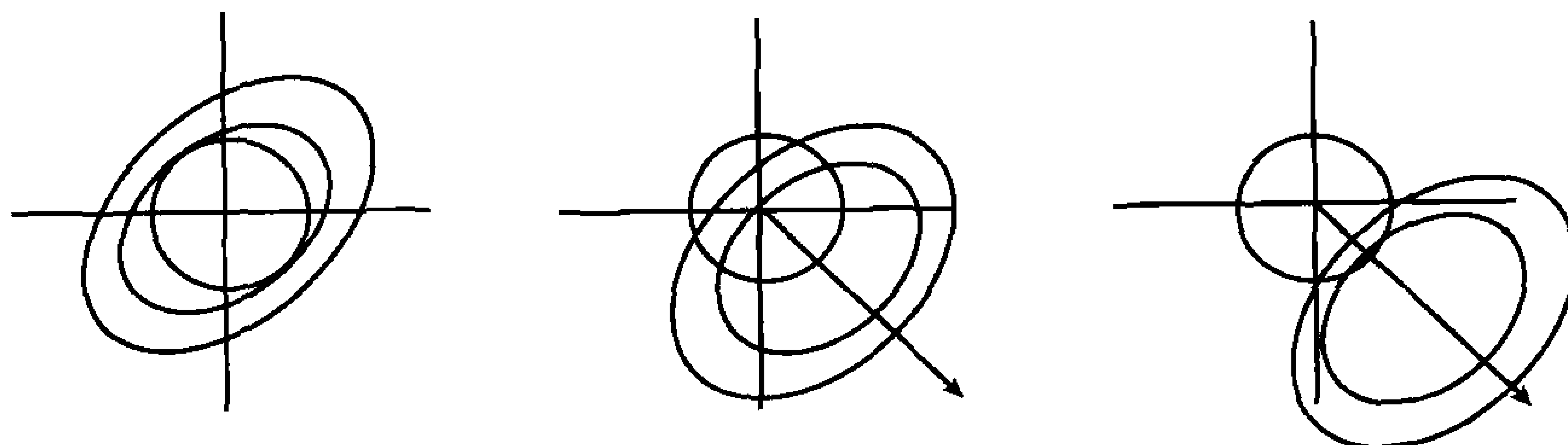


图 8.7 接受区域

定理 8.10.5 设 E 是 n 维空间中的一个凸集且关于原点对称. 设函数 $f(\boldsymbol{x}) \geqslant 0$ 满足 (i) $f(\boldsymbol{x}) = f(-\boldsymbol{x})$, (ii) 对于每个 $u(0 < u < \infty)$, $\{\boldsymbol{x}|f(\boldsymbol{x}) \geqslant u\} = K_u$ 是凸

的, (iii) $\int_E f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} < \infty$. 那么当 $0 \leqslant k \leqslant 1$ 时,

$$\int_E f(\boldsymbol{x}+k\boldsymbol{y})\mathrm{d}\boldsymbol{x} \geqslant \int_E f(\boldsymbol{x}+\boldsymbol{y})\mathrm{d}\boldsymbol{x}. \tag{37}$$

定理 8.10.5 基于下面的引理.

引理 8.10.12 如果 E, F 是凸的, 且关于原点对称. 那么

$$V\{(E+k\boldsymbol{y})\cap F\} \geqslant V\{(E+\boldsymbol{y})\cap F\}, \tag{38}$$

其中 $0 \leqslant k \leqslant 1$, V 表示 n 维体积.

证明 考虑集合 $\alpha(E+\boldsymbol{y})+(1-\alpha)(E-\boldsymbol{y}) = \alpha E+(1-\alpha)E+(2\alpha-1)\boldsymbol{y}$, 它的元素是点 $\alpha(\boldsymbol{x}+\boldsymbol{y})+(1-\alpha)(\boldsymbol{z}-\boldsymbol{y})$, $\boldsymbol{x}, \boldsymbol{z} \in E$. 设 $\alpha_0 = (k+1)/2$, 则 $2\alpha_0 - 1 = k$. 由 E 的凸性可得

$$\alpha_0(E+\boldsymbol{y})+(1-\alpha_0)(E-\boldsymbol{y}) \subset E+k\boldsymbol{y}. \tag{39}$$

再利用 F 的凸性可得

$$\alpha_0[(E+\boldsymbol{y})\cap F]+(1-\alpha_0)[(E-\boldsymbol{y})\cap F] \subset (E+k\boldsymbol{y})\cap F,$$

$$V\{\alpha_0[(E+\boldsymbol{y})\cap F]+(1-\alpha_0)[(E-\boldsymbol{y})\cap F]\} \leqslant V\{(E+k\boldsymbol{y})\cap F\}. \tag{40}$$

利用 Brunn-Minkowski 不等式 [例如 Bonnesen and Fenchel(1948) 中的 48 节] 可得

$$\begin{aligned} &V^{1/n}\{\alpha_0[(E+\boldsymbol{y})\cap F]+(1-\alpha_0)[(E-\boldsymbol{y})\cap F]\} \\ \geqslant &\alpha_0 V^{1/n}\{(E+\boldsymbol{y})\cap F\}+(1-\alpha_0)V^{1/n}\{(E-\boldsymbol{y})\cap F\} \\ = &\alpha_0 V^{1/n}\{(E+\boldsymbol{y})\cap F\}+(1-\alpha_0)V^{1/n}\{(-E+\boldsymbol{y})\cap(-F)\} \\ = &V^{1/n}\{(E+\boldsymbol{y})\cap F\}. \end{aligned} \tag{41}$$

最后一个不等式由 E 和 F 的对称性可得. ■

定理 8.10.5 的证明 令

$$H(u) = V\{(E+k\boldsymbol{y})\cap K_u\}, \tag{42}$$

$$H^*(u) = V\{(E+\boldsymbol{y})\cap K_u\}. \tag{43}$$

则

$$\begin{aligned} \int_E f(\boldsymbol{x}+\boldsymbol{y})\mathrm{d}\boldsymbol{x} &= \int_{E+\boldsymbol{y}} f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &= \int_{E+\boldsymbol{y}} \int_0^\infty \boldsymbol{I}_{\{0\leqslant u\leqslant f(\boldsymbol{x})\}}(u)\mathrm{d}u\mathrm{d}\boldsymbol{x} \\ &= \int_0^\infty \int_{E+\boldsymbol{y}} \boldsymbol{I}_{\{0\leqslant u\leqslant f(\boldsymbol{x})\}}(u)\mathrm{d}\boldsymbol{x}\mathrm{d}u \\ &= \int_0^\infty H^*(u)\mathrm{d}u. \end{aligned} \tag{44}$$

类似可得,

$$\int_E f(\boldsymbol{x}+k\boldsymbol{y})\mathrm{d}\boldsymbol{x}=\int_0^{\infty} H(u)\mathrm{d}u. \tag{45}$$

由引理 8.10.12 知, $H(u)\geqslant H^*(u)$. 再由 (44) 和 (45) 可得定理 8.10.5. ■

我们从 8.10.1 节中给出的典范型开始, 把问题如下简化. 设 $t=\min(m,p)$, $\nu_1,\cdots,\nu_t(\nu_1\geqslant\nu_2\geqslant\cdots\geqslant\nu_t)$ 为 $\boldsymbol{\Xi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}$ 的非零特征根, 其中 $\boldsymbol{\Xi}=\mathrm{E}(\boldsymbol{X})$.

引理 8.10.13 存在矩阵 $\boldsymbol{B}\,(p\times p)$ 和 $\boldsymbol{F}\,(m\times m)$ 满足

$$\begin{aligned}&\boldsymbol{B\Sigma B}'=\boldsymbol{I}_p, \qquad \boldsymbol{FF}'=\boldsymbol{I}_m,\\ &\boldsymbol{B\Xi F}'=\begin{cases}(\boldsymbol{D}_\nu^{\frac12},\boldsymbol{0}), & p\leqslant m,\\ \begin{pmatrix}\boldsymbol{D}_\nu^{\frac12}\\ \boldsymbol{0}\end{pmatrix}, & p>m,\end{cases}\end{aligned} \tag{46}$$

其中 $\boldsymbol{D}_\nu=\mathrm{diag}(\nu_1,\cdots,\nu_t)$.

证明 下面只证明 $p\leqslant m$ 和 $\nu_p>0$ 的情况. 其余情况类似可证. 利用附录中的定理 A.2.2 可知, 存在矩阵$\boldsymbol{B}$满足

$$\boldsymbol{B\Sigma B}'=\boldsymbol{I}, \qquad \boldsymbol{B\Xi\Xi}'\boldsymbol{B}'=\boldsymbol{D}_\nu. \tag{47}$$

设

$$\boldsymbol{F}_1=\boldsymbol{D}_\nu^{-\frac12}\boldsymbol{B\Xi}, \quad p\times m. \tag{48}$$

则

$$\boldsymbol{F}_1\boldsymbol{F}_1'=\boldsymbol{I}_p. \tag{49}$$

令 $\boldsymbol{F}'=(\boldsymbol{F}_1',\boldsymbol{F}_2')$ 为 $m\times m$ 的正交矩阵. 则有

$$\boldsymbol{B\Xi F}_2'=\boldsymbol{D}_\nu^{\frac12}\boldsymbol{F}_1\boldsymbol{F}_2'=\boldsymbol{0}, \tag{50}$$

$$\boldsymbol{B\Xi F}'=\boldsymbol{B\Xi}(\boldsymbol{F}_1',\boldsymbol{F}_2')=\boldsymbol{B\Xi}(\boldsymbol{\Xi}'\boldsymbol{B}'\boldsymbol{D}_\nu^{-\frac12},\boldsymbol{F}_2')=(\boldsymbol{D}_\nu^{\frac12},\boldsymbol{0}). \quad ■ \tag{51}$$

现在我们设

$$\boldsymbol{U}=\boldsymbol{BXF}', \qquad \boldsymbol{V}=\boldsymbol{BZ}. \tag{52}$$

则 $\boldsymbol{U},\boldsymbol{V}$ 的列独立地服从协方差阵为$\boldsymbol{I}$的正态分布, 当 $p\leqslant m$ 时, 它们的均值分别为

$$\begin{aligned}&\mathrm{E}(\boldsymbol{U})=(\boldsymbol{D}_\nu^{\frac12},\boldsymbol{0}),\\ &\mathrm{E}(\boldsymbol{V})=\boldsymbol{0}.\end{aligned} \tag{53}$$

不变检验是基于 $\boldsymbol{U}'(\boldsymbol{VV}')^{-1}\boldsymbol{U}$ 的特征根 $l_1,\cdots,l_t\,(l_1\geqslant\cdots\geqslant l_t)$ 给出的. 考虑到容许性, 我们利用 $\boldsymbol{U}'(\boldsymbol{UU}'+\boldsymbol{VV}')^{-1}\boldsymbol{U}$ 的特征根而不是 $l_i=\lambda_i/(1-\lambda_i)$. 这里使用 l_i 更自然, 因为它相应于参数值 ν_i. 下面的定理由 Das Gupta, Anderson, and Mudholkar (1964) 给出.

定理 8.10.6 若对任一组 $\boldsymbol{V}$ 的不变值和 $\boldsymbol{U}$ 的其他列向量而言, 某不变检验的接受区域在 $\boldsymbol{U}$ 任一列向量的空间中是凸的, 则该检验的功效关于每个 ν_i 单调递增.

证明 因为 $\boldsymbol{U}$ 的每一列都乘以 -1 时 $\boldsymbol{U}\boldsymbol{U}'$ 是不变的, 所以在$\boldsymbol{U}$的每个列向量中接受域关于原点对称. 设 $\boldsymbol{U}=(u_{ij}), \boldsymbol{V}=(v_{ij})$ 的密度为

$$
\begin{aligned}
&f(\boldsymbol{U},\boldsymbol{V}) \\
&=(2\pi)^{-\frac{1}{2}(n+m)p}\exp\left[-\frac{1}{2}\left\{\mathrm{tr}\boldsymbol{V}\boldsymbol{V}'+\sum_{i=1}^{t}(u_{ii}-\sqrt{\nu_i})^2+\sum_{\substack{i=1\\ j\neq i}}^{p}\sum_{j=1}^{m}u_{ij}^2\right\}\right].
\end{aligned}
\tag{54}
$$

应用定理 8.10.5 到 (54) 可知, 功效关于每个 $\sqrt{\nu_i}$ 单调递增. ■

因为凸集的子集也是凸的, 从而我们有下面的推论.

推论 8.10.3 *若一个不变检验的接受区域 A 在 $\boldsymbol{U}$ 中对于每个固定的$\boldsymbol{V}$是凸的, 则检验的功效关于每个 ν_i 单调递增.*

由此可知, Roy 最大根检验 $A: l_1\leqslant K$ 以及 Lawley-Hotelling 迹检验 A : $\mathrm{tr}\boldsymbol{U}'(\boldsymbol{V}\boldsymbol{V}')^{-1}\boldsymbol{U}$ 的功效函数关于每个 ν_i 单调递增.

为了说明似然比检验

$$A:\prod_{i=1}^{t}(1+l_i)\leqslant K \tag{55}$$

的接受区域满足定理 8.10.6 中的条件, 令

$$
\begin{aligned}
(\boldsymbol{V}\boldsymbol{V}')^{-1}&=\boldsymbol{T}'\boldsymbol{T}, \qquad \boldsymbol{T}:p\times p,\\
\boldsymbol{U}^*&=(\boldsymbol{u}_1^*,\cdots,\boldsymbol{u}_m^*)=\boldsymbol{T}\boldsymbol{U}.
\end{aligned}
\tag{56}
$$

则

$$
\begin{aligned}
\prod_{i=1}^{t}(1+l_i)&=|\boldsymbol{U}'(\boldsymbol{V}\boldsymbol{V}')^{-1}\boldsymbol{U}+\boldsymbol{I}|+|\boldsymbol{U}^{*'}\boldsymbol{U}^*+\boldsymbol{I}| \\
&=|\boldsymbol{U}^*\boldsymbol{U}^{*'}+\boldsymbol{I}|=|\boldsymbol{u}_1^*\boldsymbol{u}_1^{*'}+\boldsymbol{B}|\\
&=(\boldsymbol{u}_1^{*'}\boldsymbol{B}^{-1}\boldsymbol{u}_1^{*'}+1)|\boldsymbol{B}|\\
&=(\boldsymbol{u}_1'\boldsymbol{T}'\boldsymbol{B}^{-1}\boldsymbol{T}\boldsymbol{u}_1+1)|\boldsymbol{B}|,
\end{aligned}
\tag{57}
$$

其中 $\boldsymbol{B}=\boldsymbol{u}_2^*\boldsymbol{u}_2^{*'}+\cdots+\boldsymbol{u}_m^*\boldsymbol{u}_m^{*'}+\boldsymbol{I}$. 因为 $\boldsymbol{T}'\boldsymbol{B}^{-1}\boldsymbol{T}$ 是正定的, 所以 (55) 在 $\boldsymbol{u}_1$ 中是凸的. 因此似然比检验的功效函数关于每个 ν_i 单调递增.

当 $K<1$ 时, Bartlett-Nanda-Pillai 迹检验

$$A:\mathrm{tr}\boldsymbol{U}'(\boldsymbol{U}\boldsymbol{U}'+\boldsymbol{V}\boldsymbol{V}')^{-1}\boldsymbol{U}=\sum_{i=1}^{t}\frac{l_1}{1+l_i}\leqslant K \tag{58}$$

的接受区域是一个椭球面, 当 $K\leqslant 1$ 时, 该区域在$\boldsymbol{U}$的每一列 $\boldsymbol{u}_i$ 中是凸的.(见习题 8.36.) 当 $K>1$ 时, (58) 可能不是在$\boldsymbol{U}$的每一列中都是凸的. 作为例子读者可以计算一下 $p=2$ 时的情况.

Eaton and Perlman(1974) 证明了, 如果不变检验在$\boldsymbol{U}$和 $\boldsymbol{W}=\boldsymbol{V}\boldsymbol{V}'$ 中是凸的, 当 $(\sqrt{\nu_1},\cdots,\sqrt{\nu_t})\prec_w(\sqrt{\nu_1^0},\cdots,\sqrt{\nu_t^0})$ 时, 在点 $(\nu_1^0,\cdots,\nu_t^0)$ 的功效大于在点

$(\nu_1,\cdots,\nu_t)$ 的功效. 我们对该结果不做证明. Roy 最大根检验和 Lawley-Hotelling 迹检验满足这个条件, 但是似然比检验和 Bartlett-Nanda-Pillai 迹检验不满足这个条件.

Takemura 证明了若接受区域在$\boldsymbol{U}$和$\boldsymbol{W}$中是凸的, 并且 $\sqrt{\nu_1},\cdots,\sqrt{\nu_t}$ 组成的集合的功效不大于一个常量, 那么该集合是单调且凸的.

这就启发我们考虑功效函数的周线 $\Pi(\sqrt{\nu_1},\cdots,\sqrt{\nu_t})$. 定理 8.10.6 并不排除图 8.8 中的情况 (a), 类似地, Eaton-Perlman 的结果也不排除情况 (b). 最后一个结果保证了 Roy 最大根检验和 Lawley-Hotelling 迹检验的周线看起来像 (c). 这是因为当几乎没有 $\boldsymbol{\nu}_i$ 远离零时, 这两个检验更倾向于探查备择假设. 相反地, 似然比检验和 Bartlett-Nanda-Pillai 迹检验对整体偏离原假设更敏感. 必须注意到, $\sqrt{\nu}$ 空间中的凸集变换到 ν 空间中时不一定是凸的.

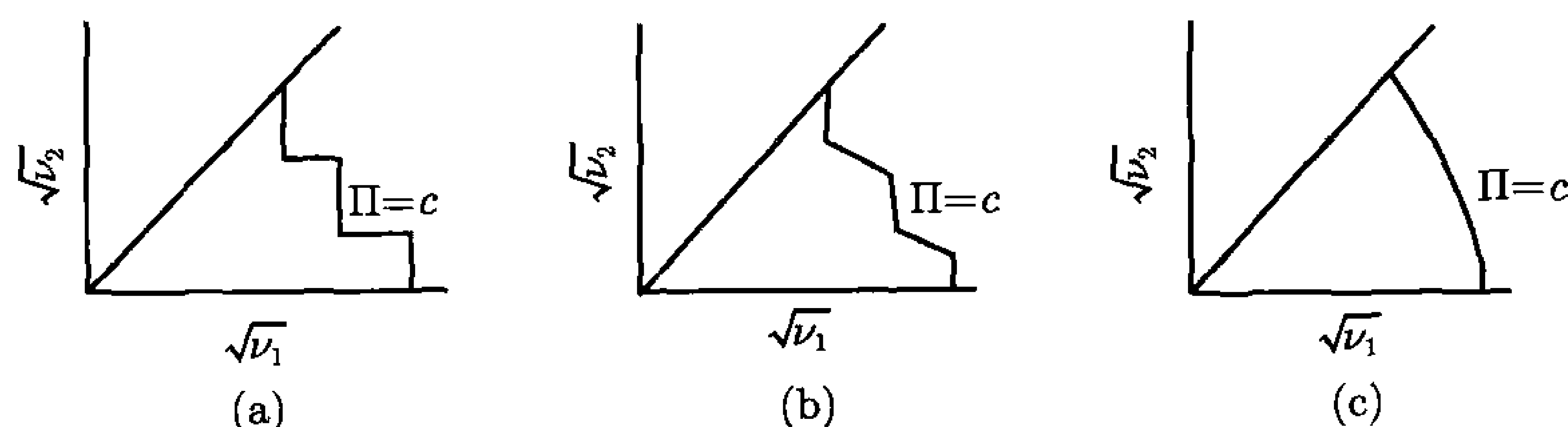


图 8.8 功效函数的周线

l_i 的非中心密度依赖于参数值 $\nu_1,\cdots,\nu_t$, Perlman and Olkin(1980) 利用该密度证明了任意一个具有单调接受区域 (在根空间中) 的不变检验都是无偏的. 该结果包含了前面提到的所有标准检验.

8.11 椭球等高分布

8.11.1 椭球等高分布的观测

8.2 节中的回归模型可以写为

$$\boldsymbol{x}_\alpha = \boldsymbol{\beta} \boldsymbol{z}_\alpha + \boldsymbol{e}_\alpha, \quad \alpha = 1,\cdots,N, \tag{1}$$

其中 $\boldsymbol{e}_\alpha$ 是满足 $\mathrm{E}(\boldsymbol{e}_\alpha)=\boldsymbol{0}$ 和 $\mathrm{E}(\boldsymbol{e}_\alpha\boldsymbol{e}_\alpha')=\boldsymbol{\Sigma}$ 的无法观测到的扰动. 我们假设 $\boldsymbol{e}_\alpha$ 的密度为 $|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g(\boldsymbol{e}'\boldsymbol{\Lambda}^{-1}\boldsymbol{e})$, 则 $\boldsymbol{\Sigma}=(\mathrm{E}(R^2)/p)\boldsymbol{\Lambda}$, 其中 $R^2=\boldsymbol{e}_\alpha'\boldsymbol{\Lambda}^{-1}\boldsymbol{e}_\alpha$. 一般来说, 很难求得或简洁地表达出 $\boldsymbol{B}=\sum_{\alpha=1}^N \boldsymbol{x}_\alpha\boldsymbol{z}_\alpha'\boldsymbol{A}^{-1}$ 和 $N\boldsymbol{\Sigma}=\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)'$ 的精确分布. 尽管如此, 因为 $\boldsymbol{B}$ 的期望值为 $\boldsymbol{\beta}$, $\mathrm{vec}\boldsymbol{B}$ 协方差阵为 $\boldsymbol{\Sigma}\otimes\boldsymbol{A}^{-1}$, 其中 $\boldsymbol{A}=\sum_{\alpha=1}^N \boldsymbol{z}_\alpha\boldsymbol{z}_\alpha'$, 我们可以探索 $\boldsymbol{B}$ 和 $N\hat{\boldsymbol{\Sigma}}$ 的大样本分布.

定理 8.11.1 *假设 $(1/N)\boldsymbol{A}\to\boldsymbol{A}_0$, $\boldsymbol{z}_\alpha'\boldsymbol{z}_\alpha <$ 常量, $\alpha=1,\cdots,N$, 而且 $\boldsymbol{e}_\alpha$ 要么独立同分布, 要么独立且对于某个 $\varepsilon>0$ 有 $\mathrm{E}|\boldsymbol{e}_\alpha'\boldsymbol{e}_\alpha|^{2+\varepsilon}<$ 常量. 那么 $\boldsymbol{B}\xrightarrow{p}\boldsymbol{\beta}$, 且*

$\sqrt{N}\mathrm{vec}(\boldsymbol{B}-\boldsymbol{\beta})$ 的极限分布是均值为0, 协方差阵为 $\boldsymbol{\Sigma}\otimes\boldsymbol{A}_0^{-1}$ 的正态分布.

Anderson(1971) 中的定理 5.5.13 就是定理 8.11.1. 在文献中有许多其他的替代假设. 在它的假设下, $\hat{\boldsymbol{\Sigma}}_\Omega\xrightarrow{p}\boldsymbol{\Sigma}$. 该结果建立了关于 $\boldsymbol{\beta}$ 的原假设检验准则的大样本理论.

考虑检验原假设

$$H:\boldsymbol{\beta}=\boldsymbol{\beta}^*, \tag{2}$$

其中 $\boldsymbol{\beta}^*$ 是给定的. 8.3 节通过把 $\boldsymbol{\beta}$ 分块为 $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)$, 给出了关于 $\boldsymbol{\beta}$ 的更一般的假设. 像 8.3 节那样, 通过变换 (4), 假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_1^*$ 可以变换为上面 (1) 中假设的形式.

设

$$\boldsymbol{G}=\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha-\boldsymbol{B}\boldsymbol{z}_\alpha)'=N\hat{\boldsymbol{\Sigma}}_\Omega, \tag{3}$$

$$\boldsymbol{H}=(\boldsymbol{B}-\boldsymbol{\beta})\boldsymbol{A}(\boldsymbol{B}-\boldsymbol{\beta})'. \tag{4}$$

引理 8.11.1 *在定理 8.11.1 的条件下, $\boldsymbol{H}$的极限分布为 $W(\boldsymbol{\Sigma},q)$.*

证明 把$\boldsymbol{H}$表示为

$$\boldsymbol{H}=\sqrt{N}(\boldsymbol{B}-\boldsymbol{\beta})\frac{1}{N}\boldsymbol{A}\sqrt{N}(\boldsymbol{B}-\boldsymbol{\beta})'. \tag{5}$$

则由定理 8.11.1 和 8.4 节中的 (4) 可得该引理. ■

我们可以把似然比准则表示为

$$\begin{aligned}-2\ln\lambda&=-N\ln U=N\ln|\boldsymbol{I}+\boldsymbol{G}^{-1}\boldsymbol{H}| \\ &=N\ln\left|\boldsymbol{I}+\frac{1}{N}(\frac{1}{N}\boldsymbol{G})^{-1}\boldsymbol{H}\right|.\end{aligned} \tag{6}$$

定理 8.11.2 *在定理 8.11.1 的条件下, 若原假设为真, 则*

$$-2\ln\lambda\xrightarrow{d}\chi^2_{pq}. \tag{7}$$

证明 因为 $|\boldsymbol{I}+x\boldsymbol{C}|=1+x\mathrm{tr}\boldsymbol{C}+O(x^2)$(定理 A.4.8), 所以当 $N\to\infty$ 时, $N\ln|\boldsymbol{I}+N^{-1}\boldsymbol{C}|=\mathrm{tr}\boldsymbol{C}+O_p(N^{-1})$.

因为 $(1/N)\boldsymbol{G}\xrightarrow{p}\boldsymbol{\Sigma},(1/N)\boldsymbol{A}\to\boldsymbol{A}_0$, 以及 $\sqrt{N}\mathrm{vec}(\boldsymbol{B}'-\boldsymbol{\beta}')$ 的极限分布为 $N(\boldsymbol{0},\boldsymbol{\Sigma}\otimes\boldsymbol{A}_0^{-1})$, 所以

$$\begin{aligned}\mathrm{tr}\left(\frac{1}{N}\boldsymbol{G}\right)^{-1}\boldsymbol{H}&=N\sum_{i,j=1}^{p}\sum_{g,h=1}^{q}g^{ij}(b_{ig}-\beta_{ig})a_{gh}(b_{jh}-\beta_{jh}) \\ &=[\mathrm{vec}(\boldsymbol{B}'-\boldsymbol{\beta}')]'\left(\frac{1}{N}\boldsymbol{G}^{-1}\otimes\boldsymbol{A}\right)\mathrm{vec}(\boldsymbol{B}'-\boldsymbol{\beta}')\xrightarrow{d}\chi^2_{pq}.\end{aligned} \tag{8}$$

■

对于来自于正态分布的样本, 定理 8.11.2 与定理 8.5.2 给出的 $-2\ln\lambda$ 的渐近展开的第一项是一致的. 利过这个 χ^2 分布就可以使用 8.3 节和 8.4 节所讨论的检验和置信过程.

准则 $U = \lambda^{2/N}$ 可以写为 $U = \prod_{i=1}^{p} V_i$, 其中 V_i 由 8.4 节的 (8) 式定义. V_i 具有 U 的形式, 也就是说它是 $x_{i\alpha}$ 关于 $x_{1\alpha}, \cdots, x_{i-1,\alpha}, \boldsymbol{z}_\alpha$ 回归的残差平方和与关于 $x_{1\alpha}, \cdots, x_{i-1,\alpha}$ 回归的残差平方和之比. 所以在原假设下, $V_1, \cdots, V_p$ 渐近独立而且 $-N \ln V_i \xrightarrow{d} \chi_q^2$. 因此 $-N \ln U = -N \sum_{i=1}^{p} \ln V_i \xrightarrow{d} \chi_{pq}^2$. 该结论验证了逐步下降法的渐近性.

8.6 节给出了一般线性假设的其他几个准则: Lawley-Hotelling 迹准则 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$, Bartlett-Nanda-Pillai 迹准则 $\mathrm{tr}\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$ 和 Roy 最大根准则 $\boldsymbol{H}\boldsymbol{G}^{-1}$ 或 $\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$. $N\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 和 $N\mathrm{tr}\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$ 的极限分布也是 χ_{pq}^2. $N\boldsymbol{H}\boldsymbol{G}^{-1}$ 或 $N\boldsymbol{H}(\boldsymbol{G}+\boldsymbol{H})^{-1}$ 的最大特征根的极限分布与$\boldsymbol{H}$的最大特征根的分布都是 $W(\boldsymbol{I}, q)$(引理 8.11.1). 这些准则的分位数可以在附录 B 中找到.

8.11.2 椭球等高矩阵的分布

在 8.3.2 节, 来自于因变量的 $p \times N$ 观测阵定义为 $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$, 来自于自变量的 $q \times N$ 观测阵定义为 $\boldsymbol{Z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N)$; 这两个矩阵通过等式 $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\beta}\boldsymbol{Z}$ 相联系. 注意在本章的观测阵有 N 列而不是 N 行.

设 $\boldsymbol{E} = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_N)$ 为一个 $p \times N$ 随机阵, 密度为 $|\boldsymbol{\Lambda}|^{-N/2} g[\boldsymbol{F}^{-1}\boldsymbol{E}\boldsymbol{E}'(\boldsymbol{F}')^{-1}]$, 其中 $\boldsymbol{\Lambda} = \boldsymbol{F}\boldsymbol{F}'$. 定义$\boldsymbol{X}$满足

$$\boldsymbol{X} = \boldsymbol{\beta}\boldsymbol{Z} + \boldsymbol{E}. \tag{9}$$

在这些项中 $\boldsymbol{\beta}$ 的最小二乘估计为

$$\boldsymbol{B} = \boldsymbol{X}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1} = \boldsymbol{C}\boldsymbol{A}^{-1}, \tag{10}$$

其中 $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{Z}' = \sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{z}_\alpha'$ 和 $\boldsymbol{A} = \boldsymbol{Z}'\boldsymbol{Z} = \sum_{\alpha=1}^{N} \boldsymbol{z}_\alpha \boldsymbol{z}_\alpha'$. 注意当$\boldsymbol{E}$右乘一个 $N \times N$ 的正交矩阵时, $\boldsymbol{E}$的密度保持不变; 即 $\boldsymbol{E}'$ 是左球面的. 因此 $\boldsymbol{E}'$ 有如下随机表示

$$\boldsymbol{E}' \stackrel{d}{=} \boldsymbol{U}\boldsymbol{T}\boldsymbol{F}', \tag{11}$$

其中$\boldsymbol{U}$是 $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_p$ 上的均匀分布, $\boldsymbol{T}$是具有非负对角元的下三角矩阵, 且满足 $\boldsymbol{E}\boldsymbol{E}' = \boldsymbol{T}\boldsymbol{T}'$, $\boldsymbol{F}$是具有非负对角元的下三角矩阵, 且满足 $\boldsymbol{F}\boldsymbol{F}' = \boldsymbol{\Sigma}$. 我们可以写为

$$\boldsymbol{B} - \boldsymbol{\beta} = \boldsymbol{E}\boldsymbol{Z}'\boldsymbol{A}^{-1} \stackrel{d}{=} \boldsymbol{F}\boldsymbol{T}'\boldsymbol{U}'\boldsymbol{Z}'\boldsymbol{A}^{-1}, \tag{12}$$

$$\boldsymbol{H} = (\boldsymbol{B} - \boldsymbol{\beta})\boldsymbol{A}(\boldsymbol{B} - \boldsymbol{\beta})' = \boldsymbol{E}\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}\boldsymbol{E}' \stackrel{d}{=} \boldsymbol{F}\boldsymbol{T}'\boldsymbol{U}'(\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z})\boldsymbol{U}\boldsymbol{T}\boldsymbol{F}', \tag{13}$$

$$\begin{aligned} \boldsymbol{G} &= (\boldsymbol{X} - \boldsymbol{\beta}\boldsymbol{Z})(\boldsymbol{X} - \boldsymbol{\beta}\boldsymbol{Z})' - \boldsymbol{H} = \boldsymbol{E}\boldsymbol{E}' - \boldsymbol{H} \\ &= \boldsymbol{E}(\boldsymbol{I}_N - \boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z})\boldsymbol{E}' = \boldsymbol{F}\boldsymbol{T}'\boldsymbol{U}'(\boldsymbol{I}_N - \boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z})\boldsymbol{U}\boldsymbol{T}\boldsymbol{F}'. \end{aligned} \tag{14}$$

8.6 节已经证明了, 对于 $H : \boldsymbol{\beta} = 0$ 的似然比准则, Lawley-Hotelling 迹准则, Bartlett-Nanda-Pillai 迹准则和 Roy 最大根准则关于线性变换 $\boldsymbol{x} \to \boldsymbol{K}\boldsymbol{x}$ 保持不变. 所以由推论 4.5.5 可得下面的定理.

定理 8.11.3 *如果原假设为 $\boldsymbol{\beta}=\mathbf{0}$, 当 $\boldsymbol{E}'$ 的分布为左球面的时, 每个不变准则的分布与在正态下的分布相同.*

因此 8.7 节中的检验和置信区域对左球面分布 $\boldsymbol{E}'$ 也成立.

$\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}$ 和 $\boldsymbol{I}_N-\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}$ 分别是秩为 q 和 $N-q$ 的幂等矩阵. 存在一个正交矩阵 $\boldsymbol{O}_N$ 满足

$$\boldsymbol{O}\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z}\boldsymbol{O}'=\begin{bmatrix}\boldsymbol{I}_q & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix},\qquad \boldsymbol{O}(\boldsymbol{I}_N-\boldsymbol{Z}'\boldsymbol{A}^{-1}\boldsymbol{Z})\boldsymbol{O}'=\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{N-q}\end{bmatrix}. \tag{15}$$

变换 $\boldsymbol{V}=\boldsymbol{O}'\boldsymbol{U}$ 在 $\boldsymbol{V}'\boldsymbol{V}=\boldsymbol{I}_p$ 上是均匀分布的, 且

$$\boldsymbol{H}=\boldsymbol{K}\boldsymbol{V}'\begin{bmatrix}\boldsymbol{I}_q & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{V}\boldsymbol{K}',\qquad \boldsymbol{G}=\boldsymbol{K}\boldsymbol{V}'\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{N-q}\end{bmatrix}\boldsymbol{V}\boldsymbol{K}', \tag{16}$$

其中 $\boldsymbol{K}=\boldsymbol{F}\boldsymbol{T}'$.

例如, 迹准则 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 为

$$\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}=\boldsymbol{V}'\begin{bmatrix}\boldsymbol{I}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\boldsymbol{V}\left(\boldsymbol{V}'\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{N-q}\end{bmatrix}\boldsymbol{V}\right)^{-1}. \tag{17}$$

任一个不变准则的分布只依赖于 $\boldsymbol{U}$ (或 $\boldsymbol{V}$), 而与$\boldsymbol{T}$无关.

因为 $\boldsymbol{G}+\boldsymbol{H}=\boldsymbol{F}\boldsymbol{T}'\boldsymbol{T}\boldsymbol{F}'$ 与$\boldsymbol{U}$独立, 所以$\boldsymbol{X}$的线性变换可以基于 $\boldsymbol{G}+\boldsymbol{H}$. 设 $\boldsymbol{D}$ 是可能依赖于 $\boldsymbol{G}+\boldsymbol{H}$ 的秩为 r 的 $p\times r$ 矩阵. 定义 $\boldsymbol{x}_\alpha^*=\boldsymbol{D}'\boldsymbol{x}_\alpha$. 这时 $\infty\mathrm{E}(\boldsymbol{x}_\alpha^*)=(\boldsymbol{D}'\boldsymbol{\beta})\boldsymbol{z}_\alpha$. 由假设 $\boldsymbol{\beta}=\mathbf{0}$ 可得 $\boldsymbol{D}'\boldsymbol{\beta}=\mathbf{0}$. 设 $\boldsymbol{X}^*=(\boldsymbol{x}_1^*,\cdots,\boldsymbol{x}_N^*)=\boldsymbol{D}'\boldsymbol{X}$, $\boldsymbol{\beta}_D=\boldsymbol{D}'\boldsymbol{\beta},\boldsymbol{E}_D=\boldsymbol{D}'\boldsymbol{E},\boldsymbol{H}_D=\boldsymbol{D}'\boldsymbol{H}\boldsymbol{D},\boldsymbol{G}_D=\boldsymbol{D}'\boldsymbol{G}\boldsymbol{D}$, 则 $\boldsymbol{E}_D'=\boldsymbol{E}'\boldsymbol{D}\overset{d}{=}\boldsymbol{U}\boldsymbol{T}\boldsymbol{F}'\boldsymbol{D}'$. 所以 $\boldsymbol{\beta}_D=\mathbf{0}$ 的不变检验准则就是 $\boldsymbol{\beta}=\mathbf{0}$ 的, 且在原假设下与正态分布的情况具有相同的分布, 只是把 p 换成 r.

习 题

8.1 (8.2.2 节) 考虑下面的样本 ($N=8$):

谷物的重量	40	17	9	15	6	12	5	9
稻草的重量	53	19	10	29	13	27	19	30
施肥量	24	11	5	12	7	14	11	18

设 $z_{2\alpha}=1$, $z_{1\alpha}$ 表示在第 α 个地点的施肥量. 对该样本估计 $\boldsymbol{\beta}$. 在显著性水平 0.01 下, 检验假设 $\boldsymbol{\beta}_1=0$.

8.2 (8.2 节) 证明定理 3.2.1 是定理 8.2.1 的特殊情况. [提示: 设 $q=1,z_\alpha=1,\boldsymbol{\beta}=\mu$.]

8.3 (8.2 节) 证明定理 8.2.3.

8.4 (8.2 节) 证明 $\hat{\boldsymbol{\beta}}$ 极小化广义方差

$$\left|\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\beta}\boldsymbol{z}_\alpha)(\boldsymbol{x}_\alpha - \boldsymbol{\beta}\boldsymbol{z}_\alpha)'\right|.$$

8.5 (8.3 节) 在下面的数据 [来自 Woltz, Reid, and Colwell (1948), R. L. Anderson and Bancroft (1952) 曾经用过] 中变量分别为: x_1, 香烟的燃烧率; x_2, 尼古丁的百分比; z_1, 氮的百分比; z_2, 氯的百分比; z_3, 钾的百分比; z_4, 磷的百分比; z_5, 钙的百分比; z_6, 镁的百分比; $z_7 = 1$; $N = 25$:

$$\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha = \begin{pmatrix} 42.20 \\ 54.03 \end{pmatrix}, \quad \sum_{\alpha=1}^{N}\boldsymbol{z}_\alpha = \begin{pmatrix} 53.92 \\ 62.02 \\ 56.00 \\ 12.25 \\ 89.79 \\ 24.10 \\ 25 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = \begin{pmatrix} 0.6690 & 0.4527 \\ 0.4527 & 6.5921 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N}(\boldsymbol{z}_\alpha - \bar{\boldsymbol{z}})(\boldsymbol{z}_\alpha - \bar{\boldsymbol{z}})'$$

$$= \begin{pmatrix} 1.8311 & -0.3589 & -0.0125 & -0.0244 & 1.6379 & 0.5057 & 0 \\ -0.3589 & 8.8102 & -0.3469 & 0.0352 & 0.7920 & 0.2173 & 0 \\ -0.0125 & -0.3469 & 1.5818 & -0.0415 & -1.4278 & -0.4753 & 0 \\ -0.0244 & 0.0352 & -0.0415 & 0.0258 & 0.0043 & 0.0154 & 0 \\ 1.6379 & 0.7920 & -1.4278 & 0.0043 & 3.7248 & 0.9120 & 0 \\ 0.5057 & 0.2173 & -0.4753 & 0.0154 & 0.9120 & 0.3828 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\sum_{\alpha=1}^{N}(\boldsymbol{z}_\alpha - \bar{\boldsymbol{z}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = \begin{pmatrix} 0.2501 & 2.6691 \\ -1.5136 & -2.0617 \\ 0.5007 & -0.9503 \\ -0.0421 & -0.0187 \\ -0.1914 & 3.4020 \\ -0.1586 & 1.1663 \\ 0 & 0 \end{pmatrix},$$

(a) 估计 x_1 和 x_2 关于 z_1, z_5, z_6, z_7 的回归方程.

(b) 估计关于全部七个变量的回归方程.

(c) 检验 z_2, z_3, z_4 回归系数为 0 的假设.

8.6 (8.3 节) 设 $q=2$, $z_{1\alpha}=w_\alpha$ (标量), $z_{2\alpha}=1$. 证明假设检验 $\boldsymbol{\beta}_1=\mathbf{0}$ 的 U 统计量是 T^2 统计量的单调函数, 并给出 T^2 统计量的简单形式. (见习题 5.1.)

8.7 (8.3 节) 设 $z_{p\alpha}=1, q_2=1$ 和

$$A^*=\left[\sum_\alpha (z_{i\alpha}-\bar{z}_i)(z_{j\alpha}-\bar{z}_j)\right],\quad i,j=1,\cdots,q_1=q-1.$$

证明

$$(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)(A_{11}-A_{12}A_{22}^{-1}A_{21})(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)'=(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)A^*(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1)'.$$

8.8 (8.3 节) 令 $q_1=q_2$. 如何检验假设 $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2$?

8.9 (8.3 节) 证明

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{1\Omega}&=\sum_\alpha \boldsymbol{x}_\alpha(\boldsymbol{z}_\alpha^{(1)}-A_{12}A_{22}^{-1}\boldsymbol{z}_\alpha^{(2)})'\left[\sum_\alpha(\boldsymbol{z}_\alpha^{(1)}-A_{12}A_{22}^{-1}\boldsymbol{z}_\alpha^{(2)})(\boldsymbol{z}_\alpha^{(1)}-A_{12}A_{22}^{-1}\boldsymbol{z}_\alpha^{(2)})'\right]^{-1}\\&=(C_1-C_2A_{22}^{-1}A_{21})(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}.\end{aligned}$$

8.10 (8.4 节) 通过对比定理 8.2.2 和习题 8.9, 证明引理 8.4.1.

8.11 (8.4 节) 通过说明 $\hat{\boldsymbol{\beta}}_{1\Omega}$ 和 $\hat{\boldsymbol{\beta}}_{2\omega}$ 的密度为

$$K_1\exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)A_{11\cdot2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'\right]$$
$$\cdot K_2\exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)A_{22}(\hat{\boldsymbol{\beta}}_{2\omega}-\boldsymbol{\beta}_2)'\right]$$

证明引理 8.4.1.

8.12 (8.4 节) 证明 $U_{3,3,n}$ 的分布函数为

$$\begin{aligned}&I_u\left(\frac{1}{2}n-1,\frac{3}{2}\right)+\frac{\Gamma(n+2)\Gamma[\frac{1}{2}(n+1)]}{\Gamma(n-1)\Gamma(\frac{1}{2}n-1)\sqrt{\pi}}\\&\cdot\left\{\frac{2u^{\frac{1}{2}n-1}\sqrt{1-u}}{n(n-1)}+\frac{u^{\frac{1}{2}(n-1)}}{n-1}\left[\arcsin(2u-1)-\frac{1}{2}\pi\right]\right.\\&\left.+\frac{2u^{\frac{1}{2}n}}{n}\ln\left(\frac{1+\sqrt{1-u}}{\sqrt{u}}\right)+\frac{2u^{\frac{1}{2}n-1}(1-u)^{\frac{3}{2}}}{3(n+1)}\right\}.\end{aligned}$$

[提示: 利用定理 8.4.4. 区域 $\{0\leqslant z_1\leqslant 1, 0\leqslant z_2\leqslant 1, z_1^2z_2\leqslant u\}$ 是 $\{0\leqslant z_1\leqslant 1, 0\leqslant z_2\leqslant u\}$ 和 $\{0\leqslant z_1\leqslant u/z_2, u\leqslant z_2\leqslant 1\}$ 的并.]

8.13 (8.4 节) 求 $\Pr\{U_{4,3,n}\geqslant u\}$.

8.14 (8.4 节) 求 $\Pr\{U_{4,4,n}\geqslant u\}$.

8.15 (8.4 节) 当 $p\leqslant m$ 时, 利用 $\boldsymbol{G}$ 和 $\boldsymbol{H}$ 的密度求 $\infty\mathrm{E}(U^h)$. [提示: 利用事实: 若$\boldsymbol{K}$的密度为 $W(\boldsymbol{\Sigma},s)$, $\boldsymbol{V}_1,\cdots,\boldsymbol{V}_t$ 相互独立且都服从 $N(\mathbf{0},\boldsymbol{\Sigma})$, 则 $\boldsymbol{K}+\sum_{i=1}^t\boldsymbol{V}_i\boldsymbol{V}_i'$ 的密度为 $W(\boldsymbol{\Sigma},s+t)$.]

8.16 (8.4 节)

(a) 证明当 p 是偶数时, $\boldsymbol{Y}=\ln U_{p,m,n}$ 的特征函数 (记为 $\phi(t)=\infty\mathrm{E}(\mathrm{e}^{itY})$) 是一个多项式的倒数.

(b) 简述利用残差得到$\boldsymbol{Y}$的特征函数的逆的方法.

(c) 说明 U 的密度是关于 $\sqrt{u}$ 和 $\ln u$ 的多项式, 可能还有一个因子 $u^{-\frac{1}{2}}$.

8.17 (8.5 节) 利用分布的渐近展开计算 $\Pr\{-k\ln U_{3,3,n}\leqslant M^*\}$, 其中 n 和 M^* 分别为

(a) $n=8,\quad M^*=14.7,$ (b) $n=8,\quad M^*=21.7,$

(c) $n=16,\quad M^*=14.7,$ (d) $n=16,\quad M^*=21.7.$

(要么保留到小数点后第三位, 要么展开到 k^{-4} 项.)

8.18 (8.5 节) 分别利用(a) $-2\ln\lambda$ 作为 χ^2 和(b) $-k\ln U$作为 χ^2, 求当 $p=3, q_1=4, n=N-q=20$ 时 $k\ln U$ 的 50% 的分位数. 利用这个展开的更多项准确地计算(a)和(b)中的结果的显著性水平.

8.19 (8.6.5 节) 证明对于 $l_i\geqslant 0, i=1,\cdots,p$,

$$\sum_{i=1}^{p}\frac{l_i}{1+l_i}\leqslant \ln\prod_{i=1}^{p}(1+l_i)\leqslant\sum_{i=1}^{p}l_i.$$

评论: 不等式暗含着 Bartlett-Nanda-Pillai 迹的值、似然比准则的负对数和 Lawley-Hotelling 迹的值的一个顺序.

8.20 (8.6 节) *多元 β 密度*. 设$\boldsymbol{H}$和$\boldsymbol{G}$相互独立, 且分布分别服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)$. 设$\boldsymbol{C}$是满足 $\boldsymbol{CC}'=\boldsymbol{H}+\boldsymbol{G}$ 的矩阵, 并设

$$\boldsymbol{L}=\boldsymbol{C}^{-1}\boldsymbol{H}\boldsymbol{C}'^{-1}.$$

证明当$\boldsymbol{L}$和 $\boldsymbol{I}-\boldsymbol{L}$ 正定时$\boldsymbol{L}$的密度为

$$\frac{\Gamma_p[\frac{1}{2}(m+n)]}{\Gamma_p(\frac{1}{2}m)\Gamma_p(\frac{1}{2}n)}|\boldsymbol{L}|^{\frac{1}{2}(m-p-1)}|\boldsymbol{I}-\boldsymbol{L}|^{\frac{1}{2}(m-p-1)},$$

否则为 0.

8.21 (8.9 节) 设 $\boldsymbol{Y}_{ij}$ (p 维向量) 的分布为 $N(\boldsymbol{\mu}_{ij},\boldsymbol{\Sigma})$, 其中 $\infty\mathrm{E}(\boldsymbol{Y}_{ij})=\boldsymbol{\mu}_{ij}=\boldsymbol{\mu}+\boldsymbol{\lambda}_i+\boldsymbol{\nu}_j+\boldsymbol{\gamma}_{ij},\sum_i\boldsymbol{\lambda}_i=\boldsymbol{0}=\sum_j\boldsymbol{\nu}_j=\sum_i\boldsymbol{\gamma}_{ij}=\sum_j\boldsymbol{\gamma}_{ij}$; $\boldsymbol{\gamma}_{ij}$ 是交互效应. 若 m 个观测来自于每一个 $\boldsymbol{Y}_{ij}$ (即 $y_{ij_1},\cdots,y_{ij_m}$), 怎么检验假设 $\boldsymbol{\lambda}_i=\boldsymbol{0},i=1,\cdots,r$? 怎样检验假设 $\boldsymbol{\gamma}_{ij}=\boldsymbol{0},i=1,\cdots,r,j=1,\cdots,c$?

8.22 (8.9 节) *拉丁方*. 设 $\boldsymbol{Y}_{ij}\,(i,j=1,\cdots,r)$ 的分布为 $N(\boldsymbol{\mu}_{ij},\boldsymbol{\Sigma})$, 其中 $\infty\mathrm{E}(\boldsymbol{Y}_{ij})=\boldsymbol{\mu}_{ij}=\boldsymbol{\gamma}+\boldsymbol{\lambda}_i+\boldsymbol{\nu}_j+\boldsymbol{\mu}_k, k=j-i+1(\mathrm{mod}\ r), \sum\boldsymbol{\lambda}_i=\sum\boldsymbol{\nu}_j=\sum\boldsymbol{\mu}_k=\boldsymbol{0}$.

(a) 给出关于主效应和误差的单变量方差分析表 (包括平方和, 自由度, 均方).

(b) 给出向量情形时的表格.

(c) 说明对于向量情形如何检验假设 $\boldsymbol{\lambda}_i=\boldsymbol{0},i=1,\cdots,r$.

8.23 (8.9 节) 设 x_1 是一个过程的产量, x_2 为质量. 设 $z_1=1, z_2=\pm10^0$(相对平均温度), $z_3=\pm0.75$(某种制剂流的相对度量), 和 $z_4=\pm1.50$(另一种制剂流的相对度量). [详见 Anderson(1955a).] 对于每个可能的组合 z_2,z_3 和 z_4, 我们对 x_1 和 x_2 做三次观测. $\boldsymbol{\beta}$ 的估计为

$$\hat{\boldsymbol{\beta}}=\begin{pmatrix}58.529 & -0.3829 & -5.050. & 2.308\\ 98.675 & 0.1558 & 4.144 & -0.700\end{pmatrix};$$

$s_1 = 3.090, s_2 = 1.619$ 和 $r = -0.6632$ 可以用来计算 $\boldsymbol{S}$ 和 $\hat{\boldsymbol{\Sigma}}$.

(a) 对该种情况给出方差分析的公式.

(b) 求出温度效应的一个置信区域 (即 β_{12}, β_{22}).

(c) 检验假设: 两种制剂对产量和质量没有影响.

8.24 (8.6 节) 解释定理 8.6.1 中变换之前的项; 即 $H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^*$ 和 $z_\alpha^{(1)}$.

8.25 (8.6 节) 求 $p = 2$ 时 $\mathrm{tr}\boldsymbol{H}\boldsymbol{G}^{-1}$ 的分布函数. [提示: 利用第 13 章中给出的根的分布.]

8.26 (8.10.1 节) *属于贝叶斯过程的* Bartlett-Nanda-Pillai V *检验*. 设 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_{n+m}$ 独立且都服从正态分布, 均值为 $\infty\mathrm{E}(\boldsymbol{w}_i) = \boldsymbol{\gamma}_i, i = 1, \cdots, m, \infty\mathrm{E}(\boldsymbol{w}_i) = \boldsymbol{0}, i = m+1, \cdots, m+n$, 协方差阵为 $\boldsymbol{\Sigma}$. 设 Π_0 由 $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [\boldsymbol{0}, (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}]$ 定义, 其中 $p \times m$ 矩阵$\boldsymbol{C}$的密度与 $|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(n+m)}$ 成比例, 且 $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m)$; 设 Π_1 由 $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}, (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}]$ 定义, 其中$\boldsymbol{C}$的密度与 $|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(n+m)}\mathrm{e}^{\frac{1}{2}\mathrm{tr}\boldsymbol{C}'(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}}$ 成比例.

(a) 通过证明 $\mathrm{tr}\boldsymbol{C}'(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C} < m$ 和 $|\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(n+m)}$ 的积分是有限的来说明当 $n \geqslant p$ 时测度是有限的. [提示: 设 $\boldsymbol{C} = (\boldsymbol{c}_1, \cdots, \boldsymbol{c}_m), \boldsymbol{D}_j = \boldsymbol{I} + \sum_{i=1}^{j} \boldsymbol{c}_i\boldsymbol{c}_i' = \boldsymbol{E}_j\boldsymbol{E}_j', \boldsymbol{c}_j = \boldsymbol{E}_{j-1}\boldsymbol{d}_j, j = 1, \cdots, m(\boldsymbol{E}_0 = \boldsymbol{I})$. 证明 $|\boldsymbol{D}_j| = |\boldsymbol{D}_{j-1}|(1 + \boldsymbol{d}_j'\boldsymbol{d}_j)$, 那么 $|\boldsymbol{D}_m| = \prod_{j=1}^{m}(1 + \boldsymbol{d}_j'\boldsymbol{d}_j)$. 然后利用习题 5.15.]

(b) 证明 5.6 节中的不等式 (26) 等价于

$$\mathrm{tr}\left(\sum_{i=1}^{m+n} \boldsymbol{w}_i\boldsymbol{w}_i'\right)^{-1} \sum_{i=1}^{m} \boldsymbol{w}_i\boldsymbol{w}_i' \geqslant k.$$

因此 Bartlett-Nanda-Pillai V 检验是贝叶斯的从而是容许的.

8.27 (8.10.1 节) *属于贝叶斯过程的似然比检验*. 设 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{n+m}$ 独立地服从正态分布, 均值为 $\infty\mathrm{E}(\boldsymbol{w}_i) = \boldsymbol{\gamma}_i, i = 1, \cdots, m, \infty\mathrm{E}(\boldsymbol{w}_i) = \boldsymbol{0}, i = m+1, \cdots, m+n$, 且$n \geqslant m+p$, 协方差阵为$\boldsymbol{\Sigma}$. 设 Π_0 由 $[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [\boldsymbol{0}, (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}]$ 定义, 其中 $p \times m$ 矩阵$\boldsymbol{C}$的密度与 $|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(n+m)}$ 成比例, 且 $\boldsymbol{\Gamma}_1 = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_m)$; 设 Π_1 定义为

$$[\boldsymbol{\Gamma}_1, \boldsymbol{\Sigma}] = [(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{D}, (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}],$$

其中$\boldsymbol{D}$的 m 个列是条件独立的, 且服从均值为$\boldsymbol{0}$ 协方差阵为 $[\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}]$ 的正态分布, 且$\boldsymbol{C}$的 (边缘) 密度与

$$|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(n+m)}|\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}|^{\frac{1}{2}m}$$

成比例.

(a) 说明测度是有限的. [提示: 见习题 8.26.]

(b) 证明 5.6 节中的不等式 (26) 等价于

$$\frac{|\sum_{i=1}^{m+n} \boldsymbol{w}_i\boldsymbol{w}_i'|}{|\sum_{i=m+1}^{m+n} \boldsymbol{w}_i\boldsymbol{w}_i'|} \geqslant k.$$

因此似然比检验是贝叶斯的从而是容许的.

8.28 (8.10.1 节) 似然比检验的容许性. 证明接受区域 $|\boldsymbol{ZZ}'|/|\boldsymbol{ZZ}'+\boldsymbol{XX}'|\geqslant c$ 满足定理 8.10.1 中的条件. [提示: 接受区域可以写为 $\prod_{i=1}^{t} m_i > c$, 其中 $m_i = 1-\lambda_i$, $i=1,\cdots,t$.]

8.29 (8.10.1 节) Lawley-Hotelling 检验的容许性. 证明接受区域 $\mathrm{tr}\boldsymbol{XX}'(\boldsymbol{ZZ}')^{-1}\leqslant c$ 满足定理 8.10.1 中的条件.

8.30 (8.10.1 节) Bartlett-Nanda-Pillai 迹检验的容许性. 证明接受区域 $\mathrm{tr}\boldsymbol{X}'(\boldsymbol{ZZ}'+\boldsymbol{XX}')^{-1}\boldsymbol{X}\leqslant c$ 满足定理 8.10.1 中的条件.

8.31 (8.10.1 节) 证明: 若$\boldsymbol{A}$和$\boldsymbol{B}$是正定的, 且 $\boldsymbol{A}-\boldsymbol{B}$ 是半正定的, 则 $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ 是半正定的.

8.32 (8.10.1 节) 证明 $\tilde{A}$ 的边界的 m 测度为 0. [提示: 证明 ($\tilde{A}$ 的闭包)$\subset \tilde{A}\cup C$, 其中 $C=\{\boldsymbol{V}|\boldsymbol{U}-\boldsymbol{YY}'$是奇异的$\}$.]

8.33 (8.10.1 节) 证明: 若 $A\subset R_{\leqslant}^{m}$ 是凸的且是以优势单调的, 则 A^* 是凸的. [提示: 说明

$$(p\boldsymbol{x}+q\boldsymbol{y})_{\downarrow}\succ_w p\boldsymbol{x}_{\downarrow}+q\boldsymbol{y}_{\downarrow},$$

其中

$$\boldsymbol{z}_{\downarrow}=(z_{[1]},\cdots,z_{[m]})'\in R_{\leqslant}^{m}.]$$

8.34 (8.10.1 节) 证明 $C(\boldsymbol{\lambda})$ 是凸的. [提示: 利用习题 8.33 的结果证明, 若 $\boldsymbol{x}\prec_w\boldsymbol{\lambda}$ 和 $\boldsymbol{y}\prec_w\boldsymbol{\lambda}$, 则 $(p\boldsymbol{x}+q\boldsymbol{y})\prec_w\boldsymbol{\lambda}$.]

8.35 (8.10.1 节) 证明若 A 是单调的, 则 A^* 也是单调的. [提示: 利用

$$x_{[k]}=\max_{i_1,\cdots,i_k}\{\min(x_{i_1},\cdots,x_{i_k})\}.]$$

8.36 (8.10.2 节) Bartlett-Nanda-Pillai 迹检验的功效函数的单调性. 当 $0\leqslant K\leqslant 1$ 时, 对于某一固定的半正定矩阵$\boldsymbol{B}$和正定矩阵 $\boldsymbol{B}+\boldsymbol{W}$, 证明

$$\mathrm{tr}(\boldsymbol{uu}'+\boldsymbol{B})(\boldsymbol{uu}'+\boldsymbol{B}+\boldsymbol{W})^{-1}\leqslant K$$

在 u 内是凸的. [提示: 证明

$$\begin{aligned}&(\boldsymbol{uu}'+\boldsymbol{B}+\boldsymbol{W})^{-1}\\=&(\boldsymbol{B}+\boldsymbol{W})^{-1}-\frac{1}{1+\boldsymbol{u}'(\boldsymbol{B}+\boldsymbol{W})^{-1}\boldsymbol{u}}(\boldsymbol{B}+\boldsymbol{W})^{-1}\boldsymbol{uu}'(\boldsymbol{B}+\boldsymbol{W})^{-1}.\end{aligned}$$

所得到的关于 u 的二次型包含矩阵 $(\mathrm{tr}\boldsymbol{A})\boldsymbol{I}-\boldsymbol{A}$, 其中 $\boldsymbol{A}=(\boldsymbol{B}+\boldsymbol{W})^{-\frac{1}{2}}\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{W})^{-\frac{1}{2}}$; 通过对角化$\boldsymbol{A}$说明该矩阵是半正定的.]

8.37 (8.8 节) 设 $\boldsymbol{x}_{\alpha}^{(\nu)}\,(\alpha=1,\cdots,N_{\nu})$ 是来自于 $N(\boldsymbol{\mu}^{(\nu)}\,\boldsymbol{\Sigma}),(\boldsymbol{\nu}=1,\cdots,q)$ 的观测. 哪种准则可以用来检验假设

$$\boldsymbol{\mu}^{(\nu)}=\sum_{h=1}^{m}\gamma_h c_{h\nu}+\boldsymbol{\mu},$$

其中 $c_{h\nu}$ 是给定的数, $\gamma_{\nu},\boldsymbol{\mu}$ 是未知向量? [提示: 该假设 (均值位于距离比已知的 m 维超平面) 可以变成一般线性假设的形式.]

8.38 (8.2 节) 设 $\boldsymbol{x}_{\alpha}$ 是来自于 $N(\boldsymbol{\beta}\boldsymbol{z}_{\alpha},\boldsymbol{\Sigma})\,(\alpha=1,\cdots,N)$ 的观测. 设存在已知固定向量 γ 满足 $\boldsymbol{\beta}\gamma=0$. 怎样估计 $\boldsymbol{\beta}$?

8.39 (8.8 节) 使得 (1) 不变的关于 $\boldsymbol{y}_{\alpha}^{(i)}\,(\alpha=1,\cdots,N_i,i=1,\cdots,q)$ 的最大变换群是什么? 证明在该群下, 检验 (12) 是不变的.

第 9 章　检验变量集间的独立性

9.1　引　　言

在本章中, 我们把具有联合正态分布的 p 个变量的一个集合划分成 q 个子集, 并考察这 q 个子集是否相互独立; 这等价于检验一个子集中的每个变量和另一个子集中的每个变量不相关的假设. 对于这个假设, 我们来找似然比准则, 在原假设下这个准则的矩, 某些特殊的分布以及该分布的渐近展开.

似然比准则在集合内的线性变换下是不变的; 另外一个这样的准则也将在本章讨论. 其他的检验方法是逐步下降 (step-down) 法, 它们虽然不是变换不变的, 但是很灵活. 在两个集合的情况下, 集合间的独立性等同于一个集合对另一个集合的回归为 0; 可以利用第 8 章中的准则进行检验. 本章还将研究似然比准则的一些最优性质.

9.2　变量集独立性检验的似然比准则

设 p 元向量 $\boldsymbol{X}$ 服从 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. 我们把 $\boldsymbol{X}$ 分成 q 个子向量, 它们分别含有 $p_1,p_2,\cdots,p_q$ 个元素, 即

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\\ \vdots\\ \boldsymbol{X}^{(q)}\end{pmatrix}. \tag{1}$$

类似地, 均值向量和协方差矩阵被划分成如下形式,

$$\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\\ \vdots\\ \boldsymbol{\mu}^{(q)}\end{pmatrix}, \tag{2}$$

$$\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1q}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{21} & \cdots & \boldsymbol{\Sigma}_{2q}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{\Sigma}_{q1} & \boldsymbol{\Sigma}_{q1} & \cdots & \boldsymbol{\Sigma}_{qq}\end{pmatrix}. \tag{3}$$

我们要检验的原假设是子向量 $\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)}$ 的分布相互独立, 即 $\boldsymbol{X}$ 的密度可以分解成 $\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)}$ 的密度之积. 也就是

$$H: n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{i=1}^{q} n(\boldsymbol{x}^{(i)}|\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}). \tag{4}$$

如果 $\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)}$ 是相互独立的子向量, 则

$$\mathrm{E}(\boldsymbol{X}^{(i)}-\boldsymbol{\mu}^{(i)})(\boldsymbol{X}^{(j)}-\boldsymbol{\mu}^{(j)})' = \boldsymbol{\Sigma}_{ij} = \boldsymbol{0}, \quad i \neq j. \tag{5}$$

(见 2.4 节.) 反过来, 如果 (5) 成立, 则 (4) 也成立. 因此原假设可以等价地变为 $H: \boldsymbol{\Sigma}_{ij}=0, i \neq j$. 另一种表述是假设 $\boldsymbol{\Sigma}$ 有如下形式

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_{21} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_{qq} \end{pmatrix}. \tag{6}$$

给定 $\boldsymbol{X}$ 的 N 个观测组成的样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, 似然比准则是

$$\lambda = \frac{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma}_0)}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu},\boldsymbol{\Sigma})}, \tag{7}$$

其中

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha-\boldsymbol{\mu})} \tag{8}$$

而 $L(\boldsymbol{\mu},\boldsymbol{\Sigma}_0)$ 是当 $\boldsymbol{\Sigma}_{ij}=0\,(i\neq j)$ 时的 $L(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 并且最大值是对所有的向量 $\boldsymbol{\mu}$ 和正定矩阵 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Sigma}_0$(即 $\boldsymbol{\Sigma}_{ii}$) 求得的, 正如 5.2 节中 (6) 式所得到的,

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}} \mathrm{e}^{-\frac{1}{2}pN}, \tag{9}$$

其中

$$\hat{\boldsymbol{\Sigma}}_\Omega = \frac{1}{N}\boldsymbol{A} = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'. \tag{10}$$

在原假设下,

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma}_0) = \prod_{i=1}^{q} L_i(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}), \tag{11}$$

其中

$$L_i(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}) = \prod_{\alpha=1}^{N} \frac{1}{(2\pi)^{\frac{1}{2}p_i}|\boldsymbol{\Sigma}_{ii}|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}_\alpha^{(i)}-\boldsymbol{\mu}^{(i)})'\boldsymbol{\Sigma}_{ii}^{-1}(\boldsymbol{x}_\alpha^{(i)}-\boldsymbol{\mu}^{(i)})}. \tag{12}$$

显然

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma}_0)=\prod_{i=1}^{q}\max_{\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}} L_i(\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}) \tag{13}$$
$$=\prod_{i=1}^{q}\frac{1}{(2\pi)^{\frac{1}{2}p_iN}|\hat{\boldsymbol{\Sigma}}_{ii\omega}|^{\frac{1}{2}N}}\mathrm{e}^{-\frac{1}{2}p_iN}$$
$$=\frac{1}{(2\pi)^{\frac{1}{2}pN}\prod_{i=1}^{q}|\hat{\boldsymbol{\Sigma}}_{ii\omega}|^{\frac{1}{2}N}}\mathrm{e}^{-\frac{1}{2}pN},$$

其中

$$\hat{\boldsymbol{\Sigma}}_{ii\omega}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(i)}-\bar{\boldsymbol{x}}^{(i)})'. \tag{14}$$

如果我们把 $\boldsymbol{A}$ 和 $\hat{\boldsymbol{\Sigma}}_\Omega$ 像 $\boldsymbol{\Sigma}$ 一样分块, 就得到

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1q}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2q}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{A}_{q1} & \boldsymbol{A}_{q2} & \cdots & \boldsymbol{A}_{qq}\end{pmatrix},\quad \hat{\boldsymbol{\Sigma}}_\Omega=\begin{pmatrix}\hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} & \cdots & \hat{\boldsymbol{\Sigma}}_{1q}\\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} & \cdots & \hat{\boldsymbol{\Sigma}}_{2q}\\ \vdots & \vdots & & \vdots\\ \hat{\boldsymbol{\Sigma}}_{q1} & \hat{\boldsymbol{\Sigma}}_{q2} & \cdots & \hat{\boldsymbol{\Sigma}}_{qq}\end{pmatrix}, \tag{15}$$

可以看出 $\hat{\boldsymbol{\Sigma}}_{ii\omega}=\hat{\boldsymbol{\Sigma}}_{ii}=(1/N)\boldsymbol{A}_{ii}$.

似然比准则是

$$\lambda=\frac{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}_0} L(\boldsymbol{\mu},\boldsymbol{\Sigma}_0)}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu},\boldsymbol{\Sigma})}=\frac{|\hat{\boldsymbol{\Sigma}}_\Omega|^{\frac{1}{2}N}}{\prod_{i=1}^{q}|\hat{\boldsymbol{\Sigma}}_{ii}|^{\frac{1}{2}N}}=\frac{|\boldsymbol{A}|^{\frac{1}{2}N}}{\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|^{\frac{1}{2}N}}. \tag{16}$$

似然比检验的临界区域是

$$\lambda\leqslant\lambda(\varepsilon), \tag{17}$$

其中 $\lambda(\varepsilon)$ 是一个数, 使得当 $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0$ 时 (17) 的概率是 ε. (这个数是否能够找到还有待证明.) 令

$$V=\frac{|\boldsymbol{A}|}{\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|}. \tag{18}$$

则 $\lambda=V^{\frac{1}{2}N}$ 是 V 的单增函数. 临界区域 (17) 可以等价地写为

$$V\leqslant V(\varepsilon). \tag{19}$$

定理 9.2.1 设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是从 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 抽出的 N 个观测组成的样本, 其中 $\boldsymbol{x}_\alpha,\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 均被划分为 $p_1,\cdots,p_q$ 行 (对 $\boldsymbol{\Sigma}$ 的情形还有列), 正如 (1), (2), (3) 中所示. q 个集合相互独立的似然比准则由 (16) 式给出, 其中 $\boldsymbol{A}$ 由 (10) 式定义且按照 (15) 式来划分, 似然比检验由 (17) 式给出, 等价地由 (19) 式给出, 其中 V 由 (18) 式定义, $\lambda(\varepsilon)$ 或 $V(\varepsilon)$ 是为了使得显著性水平达到 ε 而选择的.

由于 $r_{ij}=a_{ij}/\sqrt{a_{ii}a_{jj}}$, 我们有

$$|\boldsymbol{A}| = |\boldsymbol{R}| \prod_{i=1}^{p} a_{ii}, \tag{20}$$

其中

$$\boldsymbol{R} = (r_{ij}) = \begin{pmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} & \cdots & \boldsymbol{R}_{1q} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} & \cdots & \boldsymbol{R}_{2q} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{R}_{q1} & \boldsymbol{R}_{q2} & \cdots & \boldsymbol{R}_{qq} \end{pmatrix} \tag{21}$$

并且

$$|\boldsymbol{A}_{ii}| = |\boldsymbol{R}_{ii}| \prod_{j=p_1+\cdots+p_{i-1}+1}^{p_1+\cdots+p_i} a_{jj}. \tag{22}$$

因此

$$V = \frac{|\boldsymbol{A}|}{\prod |\boldsymbol{A}_{ii}|} = \frac{|\boldsymbol{R}|}{\prod |\mathbf{R}_{ii}|}. \tag{23}$$

即 V 可以由样本相关系数来完全表示.

我们可以用广义方差来解释准则 V. 每个集合 $(x_{i1}, \cdots, x_{iN})$ 可以看作是 N 维空间的一个向量, 设 $(x_{i1} - \bar{x}_i, \cdots, x_{iN} - \bar{x}_i) = \boldsymbol{z}_i$ 是在与等角线垂直的平面上的投影. 行列式 $|\boldsymbol{A}|$ 是以 $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_p$ 为主边的超平行体的 p 维平方体积. 行列式 $|\boldsymbol{A}_{ii}|$ 是以第 i 个向量为主边的超平行体的 p_i 维平方体积. 如果每个向量集和其他集合都是正交的 (即 $\boldsymbol{R}_{ij} = 0, i \neq j$), 则平方体积 $|\boldsymbol{A}|$ 是所有 $|\boldsymbol{A}_{ii}|$ 的乘积. 例如, 若 $p = 2, p_1 = p_2 = 1$, 则上面结论就等价于一个平行四边形在其两边成直角时, 其面积就等于两边长度之积. 如果集合之间几乎正交, 则 $|\boldsymbol{A}|$ 和 $\prod |\boldsymbol{A}_{ii}|$ 几乎相等, 且 V 近似等于 1.

这个准则具有不变性. 令 $\boldsymbol{C}_i$ 是任意一个 p_i 阶的非奇异矩阵, 且令

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{C}_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{C}_q \end{pmatrix}. \tag{24}$$

令 $\boldsymbol{C}\boldsymbol{x}_\alpha + \boldsymbol{d} = \boldsymbol{x}_\alpha^*$. 则对于 $\boldsymbol{x}_\alpha^*$ 的独立性准则对于 $\boldsymbol{x}_\alpha$ 也成立. 设 $\boldsymbol{A}^* = \sum_\alpha (\boldsymbol{x}_\alpha^* - \bar{\boldsymbol{x}}^*)(\boldsymbol{x}_\alpha^* - \bar{\boldsymbol{x}}^*)'$ 被划分为子矩阵 $\boldsymbol{A}_{ij}^*$, 则

$$\begin{aligned} \boldsymbol{A}_{ij}^* &= \sum_\alpha (\boldsymbol{x}_\alpha^{*(i)} - \bar{\boldsymbol{x}}^{*(i)})(\boldsymbol{x}_\alpha^{*(j)} - \bar{\boldsymbol{x}}^{*(j)})' \\ &= \boldsymbol{C}_i \sum_\alpha (\boldsymbol{x}_\alpha^{(i)} - \bar{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_\alpha^{(j)} - \bar{\boldsymbol{x}}^{(j)})' \boldsymbol{C}_j' \\ &= \boldsymbol{C}_i \boldsymbol{A}_{ij} \boldsymbol{C}_j' \end{aligned} \tag{25}$$

并且 $A^* = CAC'$. 因此

$$V^* = \frac{|A^*|}{\prod |A_{ii}^*|} = \frac{|CAC'|}{\prod |C_i A_{ii} C_i'|} \tag{26}$$
$$= \frac{|C| \cdot |A| \cdot |C'|}{\prod |C_i| \cdot |A_{ii}| \cdot |C_i'|} = \frac{|A|}{\prod |A_{ii}|} = V,$$

这是由于 $|C| = \prod |C_i|$. 所以该检验对于集合内的线性变换是不变的.

Narain(1950) 指出基于 V 的检验是严格无偏的, 即当原假设不真时拒绝它的概率比显著性水平高. [也可参见 Daly(1940).]

9.3 当原假设为真时似然比准则的分布

9.3.1 分布的特征

我们将指出在原假设下, 准则 V 的分布是一些独立变量的乘积的分布, 每个变量服从线性假设下准则 U 的分布 (8.4 节).

令

$$V_i = \frac{\begin{vmatrix} A_{11} & \cdots & A_{1,i-1} & A_{1i} \\ \vdots & \vdots & & \vdots \\ A_{i-1,1} & \cdots & A_{i-1,i-1} & A_{i-1,i} \\ A_{i1} & \cdots & A_{i,i-1} & A_{ii} \end{vmatrix}}{\begin{vmatrix} A_{11} & \cdots & A_{1,i-1} \\ \vdots & & \vdots \\ A_{i-1,1} & \cdots & A_{i-1,i-1} \end{vmatrix} \cdot |A_{ii}|}, \qquad i = 2, \cdots, q. \tag{1}$$

则 $V = V_2 V_3 \cdots V_q$. 其中 V_i 是用来检验原假设

$$H_i : \Sigma_{i1} = 0, \cdots, \Sigma_{i,i-1} = 0 \tag{2}$$

似然比准则的 $N/2$ 次方根, 即 $X^{(i)}$ 和 $(X^{(1)'}, \cdots, X^{(i-1)'})'$ 是独立的. 原假设 H 是这些假设的交.

定理 9.3.1 当 H_i 为真时, V_i 和 $U_{p_i, \bar{p}_i, n-\bar{p}_i}$ 分布相同, 其中 $n = N - 1$, 且 $\bar{p}_i = p_1 + \cdots + p_{i-1}, i = 2, \cdots, q$.

证明 矩阵 A 和 $\sum_{\alpha=1}^n Z_\alpha Z_\alpha'$ 有相同的分布, 其中 $Z_1, \cdots, Z_n$ 是独立同分布的且服从 $N(\mathbf{0}, \Sigma)$, Z_α 被划分为 $(Z_\alpha^{(1)'}, \cdots, Z_\alpha^{(q)'})'$). 则在 $Z_\alpha^{(1)} = z_\alpha^{(1)}, \cdots, Z_\alpha^{(i-1)} = z_\alpha^{(i-1)}$ $(\alpha = 1, \cdots, n)$ 的条件下, 子向量 $Z_1^{(i)}, \cdots, Z_n^{(i)}$ 是独立的, 且 $Z_\alpha^{(i)}$ 服从正态分布, 其均值为

$$\boldsymbol{\beta}_i \begin{pmatrix} z_\alpha^{(1)} \\ \vdots \\ z_\alpha^{(i-1)} \end{pmatrix} \tag{3}$$

且协方差阵为

$$\boldsymbol{\Sigma}_{ii} - \boldsymbol{\beta}_i \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}_{i-1,1} & \cdots & \boldsymbol{\Sigma}_{i-1,i-1} \end{pmatrix} \boldsymbol{\beta}_i', \tag{4}$$

其中

$$\boldsymbol{\beta}_i = (\boldsymbol{\Sigma}_{i1} \cdots \boldsymbol{\Sigma}_{i,i-1}) \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{\Sigma}_{i-1,1} & \cdots & \boldsymbol{\Sigma}_{i-1,i-1} \end{pmatrix}^{-1}. \tag{5}$$

当不考虑原假设时, $\boldsymbol{\beta}_i$ 的估计就是 (5) 式, 其中 $\boldsymbol{\Sigma}_{jk}$ 替换为 $\boldsymbol{A}_{jk}$, (4) 式的估计是把 $\boldsymbol{\Sigma}_{jk}$ 替换为 $(1/n)\boldsymbol{A}_{jk}$ 且把 $\boldsymbol{\beta}_i$ 替换为它的估计. 在原假设 H_i 下, $\boldsymbol{\beta}_i = \mathbf{0}$ 且协方差阵 (4) 是 $\boldsymbol{\Sigma}_{ii}$, 它由 $(1/n)\boldsymbol{A}_{ii}$ 来估计. H_i 的似然比准则的 $N/2$ 次方根是

$$\frac{\left| \boldsymbol{A}_{ii} - (\boldsymbol{A}_{i1}, \cdots, \boldsymbol{A}_{i,i-1}) \begin{pmatrix} \boldsymbol{A}_{11} & \cdots, & \boldsymbol{A}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{A}_{1i} \\ \vdots \\ \boldsymbol{A}_{i-1,i} \end{pmatrix} \right|}{|\boldsymbol{A}_{ii}|} \tag{6}$$

$$= \frac{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} & \boldsymbol{A}_{1i} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} & \boldsymbol{A}_{i-1,i} \\ \boldsymbol{A}_{i1} & \cdots & \boldsymbol{A}_{i,i-1} & \boldsymbol{A}_{ii} \end{vmatrix}}{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} \end{vmatrix} \cdot |\boldsymbol{A}_{ii}|},$$

也就是 V_i. 这是 p_i 维的 U 统计量, 其中条件向量为 $\bar{p}_i$ 维的, 协方差阵的估计的自由度为 $n - \bar{p}_i$. ■

定理 9.3.2　*在原假设下 V 的分布就是 $V_2V_3\cdots V_q$ 的分布, 其中 $V_2, \cdots, V_q$ 是独立的, 且 V_i 和 $U_{p_i,\bar{p}_i,n-\bar{p}_i}$ 同分布, 其中 $\bar{p}_i = p_i + \cdots + p_{i-1}$.*

证明　从定理 9.3.1 的证明中, 我们看出 V_i 的分布就是 $U_{p_i,\bar{p}_i,n-\bar{p}_i}$ 的分布, 而 $U_{p_i,\bar{p}_i,n-\bar{p}_i}$ 不依赖于条件 $z_\alpha^{(k)}, k = 1, \cdots, i-1$, $\alpha = 1, \cdots, n$. 因此 V_i 的分布不依赖于 $V_2, \cdots, V_{i-1}$. ■

定理 9.3.3 在原假设下, V 和 $\prod_{i=2}^{q}\prod_{j=1}^{p_i} X_{ij}$ 有相同的分布, 其中所有 X_{ij} 是独立的, 且 X_{ij} 有密度 $\beta[x|\frac{1}{2}(n-\bar{p}_i+1-j), \frac{1}{2}\bar{p}_i]$.

证明 这个定理可由定理 9.3.2 和定理 8.4.1 得到. ■

9.3.2 矩

定理 9.3.4 当原假设为真时, 准则的 h 阶矩为

$$\mathrm{E}(V^h) = \prod_{i=2}^{q}\left\{\prod_{j=1}^{p_i}\frac{\Gamma[\frac{1}{2}(n-\bar{p}_i+1-j)+h]\Gamma[\frac{1}{2}(n+1-j)]}{\Gamma[\frac{1}{2}(n-\bar{p}_i+1-j)]\Gamma[\frac{1}{2}(n+1-j)+h]}\right\}. \tag{7}$$

证明 因为 $V_2,\cdots,V_q$ 是独立的, 所以

$$\mathrm{E}(V^h) = \mathrm{E}(V_2^h)\mathrm{E}(V_3^h)\cdots\mathrm{E}(V_q^h). \tag{8}$$

由定理 9.3.2 可知, $\mathrm{E}(V_i^h) = \mathrm{E}(U^h_{p_i,\bar{p}_i,n-\bar{p}_i})$. 然后利用定理 8.4.3 即可得本定理. ■

若 p_i 是偶数, 即 $p_i = 2r_i, i>1$, 则利用伽玛函数的倍量公式 $\Gamma(\alpha+\frac{1}{2})\Gamma(\alpha+1) = \sqrt{\pi}\Gamma(2\alpha+1)2^{-2\alpha}$, 我们可以简化 V 的 h 阶矩为

$$\begin{aligned}\mathrm{E}(V^h) &= \prod_{i=2}^{q}\left\{\prod_{k=1}^{r_i}\frac{\Gamma(n+1-\bar{p}_i-2k+2h)\Gamma(n+1-2k)}{\Gamma(n+1-\bar{p}_i-2k)\Gamma(n+1-2k+2h)}\right\} \\ &= \prod_{i=2}^{q}\left\{\prod_{k=1}^{r_i}B^{-1}(n+1-\bar{p}_i-2k,\bar{p}_i)\right. \\ &\qquad \left.\cdot\int_0^1 x^{n+1-\bar{p}_i-2k+2h-1}(1-x)^{\bar{p}_i-1}\mathrm{d}x\right\}.\end{aligned} \tag{9}$$

因此 V 和 $\prod_{i=2}^{q}\{\prod_{k=1}^{r_i} Y_{ik}^2\}$ 同分布, 其中 Y_{ik} 独立, 且 Y_{ik} 有密度 $\beta(y; n+1-\bar{p}_i-2k, \bar{p}_i)$.

通常, 伽玛函数的倍量公式可以用来简化矩的表达式, 正如 8.4 节所示.

9.3.3 一些特殊的分布

若 $q=2$, 则 V 与 $U_{p_2,p_1,n-p_1}$ 同分布. 一些特殊的情况已经在 8.4 节讨论过了, 且可以参考给出的文献. 对于 $p_1=p_2=p_3=1$ 和 $p_1=p_2=p_3=2$ 时的分布分别在习题 9.2 和习题 9.3 中给出. Wilks(1935) 给出了一些分布, 分别是当 $p_1=p_2=1, p_3=p-2$①; $p_1=1, p_2=p_3=2$; $p_1=1, p_2=2, p_3=3$; $p_1=1, p_2=2, p_3=4$ 以及 $p_1=p_2=2, p_3=3$ 时的情况. Consul(1967a) 处理了当 $p_1=2, p_2=3, p_3$ 为偶数时的情况.

Wald and Brookner(1941) 给出了一种方法, 可以用来得到至多有一个 p_i 为奇数时的分布. 可以看出, 在利用倍量公式简化矩后, 再利用贝塔函数乘积的积分也可以得到同样的结果.

① 在 Wilks 的公式中, $\Gamma[\frac{1}{2}(N-2-i)]$ 应该是 $\Gamma[\frac{1}{2}(n-2-i)]$.

Mathai and Saxena(1973) 给出了一般情况时的精确分布. Mathai and Katiyar (1979) 给出了当 $p=3(1)10$ 和 $n=3(1)20$ 时在显著性水平 5% 和 1% (9.4 节的 $-k\ln V$) 下的一些精确的分位数.

9.4　似然比准则的分布的渐近展开

$\lambda=V^{\frac{1}{2}N}$ 的 h 阶矩为

$$\mathrm{E}(\lambda^h)=K\frac{\prod_{i=1}^{p}\Gamma\left\{\frac{1}{2}[N(1+h)-i]\right\}}{\prod_{i=1}^{q}\left\{\prod_{j=1}^{p_i}\Gamma\left\{\frac{1}{2}[N(1+h)-j]\right\}\right\}},\tag{1}$$

其中 K 的选择是使得 $\mathrm{E}(\lambda^0)=1$. 这和 8.5 节中 (1) 式的形式相同, 其中

$$\begin{aligned}&a=p,\quad b=p,\quad x_k=\frac{N}{2},\quad \xi_k=\frac{-k}{2},\quad k=1,\cdots,p,\\&y_j=\frac{N}{2},\qquad \eta_j=\frac{-j+p_i+\cdots+p_{i-1}}{2},\\&j=p_1+\cdots+p_{i-1}+1,\cdots,p_1+\cdots+p_i,\quad i=1,\cdots,q.\end{aligned}\tag{2}$$

则 $f=\frac{1}{2}[p(p+1)-\sum p_i(p_i+1)]=\frac{1}{2}(p^2-\sum p_i^2), \beta_k=\varepsilon_j=\frac{1}{2}(1-p)N$.

为了消去展开式中第二项我们选择 ρ 为

$$\rho=1-\frac{2(p^3-\sum p_i^3)+9(p^2-\sum p_i^2)}{6N(p^2-\sum p_i^2)}.\tag{3}$$

令

$$k=\rho N=N-\frac{3}{2}-\frac{p^3-\sum p_i^3}{3(p^2-\sum p_i^2)}.\tag{4}$$

则 $\omega_2=\gamma_2/k^2$, 其中 [可以参考 Box(1949)]

$$\gamma_2=\frac{p^4-\sum p_i^4}{48}-\frac{5(p^2-\sum p_i^2)}{96}-\frac{(p^3-\sum p_i^2)^3}{72(p^2-\sum p_i^2)}.\tag{5}$$

根据 8.5 节我们得到如下展开:

$$\Pr\{-k\ln V\leqslant v\}=\Pr\left\{\chi_f^2\leqslant v\right\}+\frac{\gamma_2}{k^2}\left[\Pr\left\{\chi_{f+4}^2\leqslant v\right\}-\Pr\left\{\chi_f^2\leqslant v\right\}\right]+O(k^{-3}).\tag{6}$$

如果 $q=2$, 我们可以利用 8.5 节的结果得到展开式的更多项.

如果 $p_i=1$, 我们有

$$\begin{aligned}f&=\frac{1}{2}p(p-1),\\k&=N-\frac{2p+11}{6},\\\gamma_2&=\frac{p(p-1)}{288}(2p^2-2p-13),\\\gamma_3&=\frac{p(p-1)}{3240}(p-2)(2p-1)(p+1).\end{aligned}\tag{7}$$

其余的项由 Box(1949) 给出. 如果 $p_i = 2\,(p = 2q)$,

$$\begin{aligned} f &= 2q(q-1), \\ k &= N - \frac{4q+13}{6}, \\ \gamma_2 &= \frac{q(q-1)}{72}(8q^2 - 8q - 7). \end{aligned} \tag{8}$$

表 9.1 给出了当 $p_i = 1$ 时 (6) 式的近似阶数. 对于每种情况, v 的选择是使得第一项为 0.95.

表 9.1

p	f	v	γ_2	N	k	γ_2/k^2	第二项
4	6	12.592	$\frac{11}{24}$	15	$\frac{71}{6}$	0.0033	−0.0007
5	10	18.307	$\frac{15}{8}$	15	$\frac{69}{6}$	0.0142	−0.0021
6	15	24.996	$\frac{235}{48}$	15	$\frac{67}{6}$	0.0393	−0.0043
				16	$\frac{73}{6}$	0.0331	−0.0036

若 $q = 2$, 则 8.5.3 节和 8.5.4 节中给出的近似分布是可以利用的. [也可参见 Nagao(1973c).]

9.5 其他准则

对于 $q = 2$ 的情况, 可以利用 8.6 节中考虑的准则, 其中 $\boldsymbol{G}+\boldsymbol{H}$ 用 $\boldsymbol{A}_{11}$ 来代替, $\boldsymbol{H}$ 用 $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ 来代替, 或者 $\boldsymbol{G}+\boldsymbol{H}$ 用 $\boldsymbol{A}_{22}$ 来代替, $\boldsymbol{H}$ 用 $\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ 来代替.

独立性的原假设是 $\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0 = 0$, 其中 $\boldsymbol{\Sigma}_0$ 由 9.2 节中的 (6) 式来定义. 当 $\boldsymbol{A} - \boldsymbol{A}_0$ 的对角块元素相对于 $\boldsymbol{A}_0$ 的对角块元素来说比较大时, 一个合适的检验方法应该拒绝原假设 (其中 $\boldsymbol{A}_0$ 的对角块为 $\boldsymbol{A}_{ii}$, 其余块为 0). 令非奇异矩阵 $\boldsymbol{B}_{ii}$ 满足 $\boldsymbol{B}_{ii}\boldsymbol{A}_{ii}\boldsymbol{B}_{ii}' = \boldsymbol{I}$, 即 $\boldsymbol{A}_{ii}^{-1} = \boldsymbol{B}_{ii}'\boldsymbol{B}_{ii}$, 并且令 $\boldsymbol{B}_0$ 第 i 个对角块为 $\boldsymbol{B}_{ii}$、其余块为 $\boldsymbol{0}$ 的矩阵, 则 $\boldsymbol{B}_0\boldsymbol{A}_0\boldsymbol{B}_0' = \boldsymbol{I}$, 且

$$\boldsymbol{B}_0(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{B}_0' = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{B}_{11}\boldsymbol{A}_{12}\boldsymbol{B}_{22}' & \cdots & \boldsymbol{B}_{11}\boldsymbol{A}_{1q}\boldsymbol{B}_{qq}' \\ \boldsymbol{B}_{22}\boldsymbol{A}_{21}\boldsymbol{B}_{11}' & \boldsymbol{0} & \cdots & \boldsymbol{B}_{22}\boldsymbol{A}_{2q}\boldsymbol{B}_{qq}' \\ \vdots & \vdots & & \vdots \\ \boldsymbol{B}_{qq}\boldsymbol{A}_{q1}\boldsymbol{B}_{11}' & \boldsymbol{B}_{qq}\boldsymbol{A}_{q2}\boldsymbol{B}_{22}' & \cdots & \boldsymbol{0} \end{pmatrix}. \tag{1}$$

这个矩阵在 9.2 节作用在 $\boldsymbol{A}$ 上的变换 (24) 下是不变的. $\boldsymbol{B}_{ii}$ 的不同选择相当于对 (1) 式左乘 Q_0 右乘 Q_0', 其中 $\boldsymbol{Q}_0$ 是一个对角块为正交矩阵、非对角块为 $\boldsymbol{0}$ 的矩阵.

当 (1) 中元素在某种度量下测得的数值太大时, 检验方法应该拒绝原假设. 似然比准则是 $|\boldsymbol{B}_0(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{B}_0'+\boldsymbol{I}|=|\boldsymbol{B}_0\boldsymbol{A}\boldsymbol{B}_0'|$ 的 $N/2$ 次幂.

由 Nagao(1973a) 给出的另一种度量是

$$\begin{aligned}\frac{1}{2}\mathrm{tr}\,[\boldsymbol{B}_0(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{B}_0']^2&=\frac{1}{2}\mathrm{tr}\,[(\boldsymbol{A}-\boldsymbol{A}_0)\boldsymbol{A}_0^{-1}]^2=\frac{1}{2}\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{A}_0^{-1}-\boldsymbol{I})^2\\&=\frac{1}{2}\sum_{\substack{i,j=1\\i\neq j}}^{q}\mathrm{tr}\,\boldsymbol{A}_{ij}\boldsymbol{A}_{jj}^{-1}\boldsymbol{A}_{ji}\boldsymbol{A}_{ii}^{-1}.\end{aligned}\tag{2}$$

对于 $q=2$, 这个度量是 Bartlett-Nanda-Pillai 迹准则的平均, 其中 $\boldsymbol{G}+\boldsymbol{H}$ 由 $\boldsymbol{A}_{11}$ 来代替, $\boldsymbol{H}$ 由 $\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ 来代替, 或者 $\boldsymbol{G}+\boldsymbol{H}$ 由 $\boldsymbol{A}_{22}$ 来代替, $\boldsymbol{H}$ 由 $\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$ 来代替.

这个准则乘以 n 或 N 的极限分布是 χ^2 分布, 自由度为 $f=\frac{1}{2}(p^2-\sum_{i=1}^{q}p_i^2)$, 与 $-N\ln V$ 的自由度相同, Nagao 得到了这个分布的渐近展开式

$$\begin{aligned}&\Pr\left\{\frac{1}{2}n\mathrm{tr}\,(\boldsymbol{A}\boldsymbol{A}_0^{-1}-I)^2\leqslant x\right\}\\=&\Pr\{\chi_f^2\leqslant x\}\\&+\frac{1}{n}\left[\frac{1}{12}\left(p^3-3p\sum_{i=1}^{q}p_i^2+2\sum_{i=1}^{q}p_i^3\right)\Pr\{\chi_{f+6}^2\leqslant x\}\right.\\&+\frac{1}{8}\left(-2p^3+4p\sum_{i=1}^{q}p_i^2-2\sum_{i=1}^{q}p_i^3-p^2+\sum_{i=1}^{q}p_i^2\right)\Pr\{\chi_{f+4}^2\leqslant x\}\\&+\frac{1}{4}\left(p^3-p\sum_{i=1}^{q}p_i^2+p^2-\sum_{i=1}^{q}p_i^3\right)\Pr\{\chi_{f+2}^2\leqslant x\}\\&\left.-\frac{1}{24}\left(2p^3-2\sum_{i=1}^{q}p_i^3+3p^2-3\sum_{i=1}^{q}p_i^2\right)\Pr\{\chi_f^2\leqslant x\}\right]+O(n^{-2}).\end{aligned}\tag{3}$$

9.6　逐步下降法

9.6.1　按块逐步下降

在 9.3 节中已经指出, 似然比准则的 $N/2$ 次根 V 是 $q-1$ 个准则 $V_2,\cdots,V_q$ 的乘积, 第 i 个子准则 V_i 给出了假设 H_i[见 9.3 节 (2)] 的似然比检验, 其中 H_i 表示第 i 个子向量和其前面的 $i-1$ 个子向量独立, 在原假设 $H\,[=\cap_{i=2}^{q}H_i]$ 下, $q-1$ 个准则是独立的 (定理 9.3.2). 逐步下降检验法是, 如果

$$V_i\geqslant v_i(\varepsilon_i),\quad i=2,\cdots,q,\tag{1}$$

则接受原假设, 如果存在 i 使得 $V_i<v_i(\varepsilon_i)$ 则拒绝原假设. 这里 $v_i(\varepsilon_i)$ 是一个数, 它使得当 H_i 为真时, (1) 式的概率是 $1-\varepsilon_i$. 这个方法的显著性水平 ε 满足

$$1-\varepsilon=\prod_{i=1}^{q}(1-\varepsilon_i). \tag{2}$$

子检验可以按照次序 $2,\cdots,q$ 依次做下去. 只要有一个子检验被拒绝了, 则过程就可以终止了; 如果没有子检验被拒绝, 则接受 H. 子向量的顺序不仅要依照研究者的意见, 也要考虑检验的次序.

例如, 假设对于个体的测量可以分为生理机制测量、智力测量和情感特征测量. 我们可以检验智力与生理是独立的, 而情感则与生理和智力是独立的, 或者把它们的顺序调换一下. 另外, 我们也可以检验智力与情感是独立的, 而生理与这两个因素是独立的, 或者把顺序调换一下. 还有第三对检验方法.

在 8.6 节讨论过的检验线性假设的其他准则可以类似地用来检验成分假设 $H_2,\cdots,H_q$. 当 H_i 为真时, 其准则的分布是和 $\boldsymbol{X}_\alpha^{(1)},\cdots,\boldsymbol{X}_\alpha^{(i-1)}\,(\alpha=1,\cdots,N)$ 独立的, 因此也和 $H_2,\cdots,H_{i-1}$ 的准则独立.

9.6.2 按分量逐步下降

在 8.4.5 节中我们讨论过一个按照分量的逐步下降法, 用来检验回归系数的子矩阵是一个特定的矩阵. 利用这个方法, 我们来检验原假设 H_i, 它以下面的形式出现:

$$H_i:(\boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{i2}\cdots\boldsymbol{\Sigma}_{i,i-1})\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1,i-1}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2,i-1}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{\Sigma}_{i-1,1} & \boldsymbol{\Sigma}_{i-1,2} & \cdots & \boldsymbol{\Sigma}_{i-1,i-1}\end{pmatrix}^{-1}=\mathbf{0}, \tag{3}$$

其中 $\mathbf{0}$ 是 $p_i\times\bar{p}_i$ 的矩阵. (3) 中的矩阵由 $\boldsymbol{X}^{(i)}$ 对 $(\boldsymbol{X}^{(1)\prime},\cdots,\boldsymbol{X}^{(i-1)\prime})'$ 的回归系数组成.

对于 $i=2$, 我们依次检验 $\boldsymbol{X}_{p_1+1}$ 对 $\boldsymbol{X}^{(1)}=(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_{p_1})'$ 的回归是否为 $\mathbf{0}$, 在 $\boldsymbol{X}_{p_1+2}$ 对 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}_{p_1+1}$ 的回归中, $\boldsymbol{X}_{p_1+2}$ 对 $\boldsymbol{X}^{(1)}$ 的回归是否为 $\mathbf{0}$…… 在 $\boldsymbol{X}_{p_1+p_2}$ 对 $\boldsymbol{X}^{(1)},\boldsymbol{X}_{p_1+1},\cdots,\boldsymbol{X}_{p_1+p_2-1}$ 的回归中, $\boldsymbol{X}_{p_1+p_2}$ 对 $\boldsymbol{X}_{(1)}$ 的回归是否为 $\mathbf{0}$. 这些假设等价于 (3) 中的矩阵在 $i=2$ 时第 $1,2,\cdots,p_2$ 行分别是 $\mathbf{0}$ 向量.

令 $\boldsymbol{A}_{ii}^{(k)}$ 为 $\boldsymbol{A}_{ii}$ 的左上角的 $k\times k$ 矩阵, $\boldsymbol{A}_{ij}^{(k)}$ 由 $\boldsymbol{A}_{ij}$ 的前 k 行组成, $\boldsymbol{A}_{ji}^{(k)}$ 由 $\boldsymbol{A}_{ij}$ 的前 k 列组成, $k=1,\cdots,p_i$. 则检验 (3) 的第一行为 $\mathbf{0}$ 的准则是

$$X_{i1}=\left|\boldsymbol{A}_{ii}^{(1)}-(\boldsymbol{A}_{i1}^{(1)}\cdots\boldsymbol{A}_{i,i-1}^{(1)})\begin{pmatrix}\boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1}\\ \vdots & & \vdots\\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{A}_{1i}^{(1)}\\ \vdots\\ \boldsymbol{A}_{i-1,i}^{(1)}\end{pmatrix}\right|\div\boldsymbol{A}_{ii}^{(1)}$$

$$
= \frac{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} & \boldsymbol{A}_{1i}^{(1)} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} & \boldsymbol{A}_{i-1,i}^{(1)} \\ \boldsymbol{A}_{i1}^{(1)} & \cdots & \boldsymbol{A}_{i,i-1}^{(1)} & \boldsymbol{A}_{ii}^{(1)} \end{vmatrix}}{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} \end{vmatrix} \boldsymbol{A}_{ii}^{(1)}}. \tag{4}
$$

对于 $k=1$, 检验 (3) 中矩阵第 k 行为 $\mathbf{0}$ 的准则 [见 8.4 节 (8) 式] 是

$$
X_{ik} = \frac{\left| \boldsymbol{A}_{ii}^{(k)} - (\boldsymbol{A}_{i1}^{(k)} \cdots \boldsymbol{A}_{i,i-1}^{(k)}) \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{A}_{1i}^{(k)} \\ \vdots \\ \boldsymbol{A}_{1,i-1}^{(k)} \end{pmatrix} \right|}{\left| \boldsymbol{A}_{ii}^{(k-1)} - (\boldsymbol{A}_{i1}^{(k-1)} \cdots \boldsymbol{A}_{i,i-1}^{(k-1)}) \begin{pmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} \\ \vdots & & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{A}_{1i}^{(k-1)} \\ \vdots \\ \boldsymbol{A}_{1,i-1}^{(k-1)} \end{pmatrix} \right|}
$$

$$
\div \frac{|\boldsymbol{A}_{ii}^{(k)}|}{|\boldsymbol{A}_{ii}^{(k-1)}|}
$$

$$
= \frac{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} & \boldsymbol{A}_{1i}^{(k)} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} & \boldsymbol{A}_{i-1,i}^{(k)} \\ \boldsymbol{A}_{i1}^{(k)} & \cdots & \boldsymbol{A}_{i,i-1}^{(k)} & \boldsymbol{A}_{ii}^{(k)} \end{vmatrix}}{\begin{vmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1,i-1} & \boldsymbol{A}_{1i}^{(k-1)} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{A}_{i-1,1} & \cdots & \boldsymbol{A}_{i-1,i-1} & \boldsymbol{A}_{i-1,i}^{(k-1)} \\ \boldsymbol{A}_{i1}^{(k-1)} & \cdots & \boldsymbol{A}_{i,i-1}^{(k-1)} & \boldsymbol{A}_{ii}^{(k-1)} \end{vmatrix}} \cdot \frac{|\boldsymbol{A}_{ii}^{(k-1)}|}{|\boldsymbol{A}_{ii}^{(k)}|}, \quad k=2,\cdots,p_i,\ i=2,\cdots,q. \tag{5}
$$

在原假设下这个准则有 β 密度 $\beta[x;\frac{1}{2}(n-\bar{p}_i+1-j),\frac{1}{2}p_i]$. 对于给定的 i, 这些准则 $X_{i1},\cdots,X_{ip_i}$ 是独立的 (定理 8.4.1). 由 9.6.1 节的论述可以看出, 对不同的 i 这些集合是独立的.

逐步下降法包含一列基于 $X_{21},\cdots,X_{2p_2},X_{31},\cdots,X_{qp_q}$ 的检验. 如果

$$
\frac{1-X_{ij}}{X_{ij}} \frac{n-\bar{p}_i+1-j}{p_i} > F_{p_i,n-\bar{p}_i+1-j}(\varepsilon_{ij}), \tag{6}
$$

一个特别的分量检验会导致拒绝原假设. 显著性水平是 ε, 其中

$$1-\varepsilon=\prod_{i=2}^{q}\prod_{j=1}^{p_i}(1-\varepsilon_{ij}). \tag{7}$$

这些子向量以及每个向量内的分量的选择是由研究者来决定的.

检验 H_i 的准则 V_i 是 $V_i=\prod_{k=1}^{p_i}X_{ik}$, 而检验原假设 H 的准则是

$$V=\prod_{i=2}^{q}V_i=\prod_{i=2}^{q}\prod_{k=1}^{p_i}X_{ik}. \tag{8}$$

这些都是在定理 9.3.3 中定义过的随机变量.

9.7 例　子

下面是一个工业生产时间研究的例子 [Abruzzi(1950)]. 这个研究的目的是探讨在一个服装加工厂中不同的工人在做压制操作的不同步骤时所用的时间. 整个压制操作可以分为以下六个基本操作.

(1) 挑选和摆放衣服.

(2) 压制和重新压制短褶.

(3) 把衣服重新摆放到电烙板上.

(4) 压制长褶的四分之三.

(5) 压制长褶使其平整.

(6) 把衣服挂到衣架上.

这里 $\boldsymbol{x}_\alpha$ 表示第 α 个个体的测量向量. 分量 $x_{i\alpha}$ 是他做第 i 个基本操作时所用的时间. N 是 76. 数据 (以秒为单位) 的样本均值向量和协方差阵总结如下:

$$\bar{\boldsymbol{x}}=\begin{pmatrix}9.47\\25.56\\13.25\\31.44\\27.29\\8.80\end{pmatrix}, \tag{1}$$

$$\boldsymbol{S}=\begin{pmatrix}2.57&0.85&1.56&1.79&1.33&0.42\\0.85&37.00&3.34&13.47&7.59&0.52\\1.56&3.34&8.44&5.77&2.00&0.50\\1.79&13.47&5.77&34.01&10.50&1.77\\1.33&7.59&2.00&10.50&23.01&3.43\\0.42&0.52&0.50&1.77&3.43&4.59\end{pmatrix}. \tag{2}$$

样本的标准差是 (1.604, 6.041, 2.903, 5.832, 4.798, 2.141). 样本的自相关阵是

$$\boldsymbol{R}=\begin{pmatrix} 1.000 & 0.088 & 0.334 & 0.191 & 0.173 & 0.123 \\ 0.088 & 1.000 & 0.186 & 0.384 & 0.262 & 0.040 \\ 0.334 & 0.186 & 1.000 & 0.343 & 0.144 & 0.080 \\ 0.191 & 0.384 & 0.343 & 1.000 & 0.375 & 0.142 \\ 0.173 & 0.262 & 0.144 & 0.375 & 1.000 & 0.334 \\ 0.123 & 0.040 & 0.080 & 0.142 & 0.334 & 1.000 \end{pmatrix}. \tag{3}$$

研究者感兴趣的是检验六个变量相互独立的假设. 在时间研究中经常发生的是, 一个新操作是由基本操作以不同方式组合而成的; 新操作可能会重复用到一些基本操作而不用另外一些基本操作. 如果要研究的操作中基本操作所用的时间是独立的, 则有理由猜测在新操作中它们也是独立的. 则新操作所用时间的分布就可以用各个基本操作的均值和方差来估计了.

在这个问题中准则 V 是 $V=|\boldsymbol{R}|=0.472$. 由于样本量很大, 我们可以利用渐近理论: $k=\frac{433}{6}, f=15$, 且 $-k\ln V=54.1$. 由于自由度为 15 的 χ^2 分布在显著性水平 0.01 的分位数为 30.6, 所以结果是显著的. 我们拒绝独立的假设, 不能认为基本操作所用的时间是独立的.

9.8 两个变量集的情形

对于两个变量集的情况 ($q=2$), 随机向量 $\boldsymbol{X}$, 观测向量 $\boldsymbol{x}_\alpha$, 均值向量 $\boldsymbol{\mu}$ 以及协方差矩 $\boldsymbol{\Sigma}$ 按如下方式分块:

$$\boldsymbol{X}=\begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix},\quad \boldsymbol{x}_\alpha=\begin{pmatrix} \boldsymbol{x}_\alpha^{(1)} \\ \boldsymbol{x}_\alpha^{(2)} \end{pmatrix},\quad \boldsymbol{\mu}=\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix},\quad \boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{1}$$

独立的原假设表示 $\boldsymbol{\Sigma}_{12}=0$, 即 $\boldsymbol{\Sigma}$ 有如下形式

$$\boldsymbol{\Sigma}_0=\begin{pmatrix} \boldsymbol{\Sigma}_{11} & 0 \\ 0 & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{2}$$

检验准则是

$$V=\frac{|\boldsymbol{A}|}{|\boldsymbol{A}_{11}|\cdot|\boldsymbol{A}_{22}|}. \tag{3}$$

在 9.3 节中已经指出, 当原假设为真时, 这个准则有分布 $U_{p_1,p_2,N-1-p_2}$, 它是关于回归系数的假设检验的准则 (第 8 章). 现在我们想进一步研究两个集合独立性的假设检验和一个集合对另一个集合的回归是 0 的假设检验之间的关系.

给定 $\boldsymbol{X}_\alpha^{(2)}=\boldsymbol{x}_\alpha^{(2)}$ 下 $\boldsymbol{X}_\alpha^{(1)}$ 的条件分布是 $N[\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}_\alpha^{(2)}-\boldsymbol{\mu}^{(2)}),\boldsymbol{\Sigma}_{11.2}]=N[\boldsymbol{\beta}(\boldsymbol{x}_\alpha{}^{(2)}-\bar{\boldsymbol{x}}^{(2)})+\boldsymbol{\nu},\boldsymbol{\Sigma}_{11.2}]$, 其中 $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$, $\boldsymbol{\Sigma}_{11.2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, 且

$\boldsymbol{\nu}=\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\bar{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})$. 令 $\boldsymbol{X}_\alpha^*=\boldsymbol{X}_\alpha^{(1)}$, $\boldsymbol{z}_\alpha^{*\prime}=[(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})'1]$, $\boldsymbol{\beta}^*=(\boldsymbol{\beta}\quad\boldsymbol{\nu})$, 且 $\boldsymbol{\Sigma}^*=\boldsymbol{\Sigma}_{11.2}$. 则 $\boldsymbol{X}_\alpha^*$ 的条件分布是 $N(\boldsymbol{\beta}^*\boldsymbol{z}_\alpha^*,\boldsymbol{\Sigma}^*)$. 这正是第 8 章中研究过的分布.

原假设 $\boldsymbol{\Sigma}_{12}=\boldsymbol{0}$ 等价于原假设 $\boldsymbol{\beta}=\boldsymbol{0}$. 考虑到 $\boldsymbol{x}_\alpha^{(2)}$ 固定, 由第 8 章我们知道这个假设检验的准则 (基于似然比准则) 是

$$U=\frac{\left|\sum(\boldsymbol{x}_\alpha^*-\hat{\boldsymbol{\beta}}_\Omega^*\boldsymbol{z}_\alpha^*)(\boldsymbol{x}_\alpha^*-\hat{\boldsymbol{\beta}}_\Omega^*\boldsymbol{z}_\alpha^*)'\right|}{\left|\sum(\boldsymbol{x}_\alpha^*-\hat{\boldsymbol{\beta}}_{2\omega}^*\boldsymbol{z}_\alpha^{*(2)})(\boldsymbol{x}_\alpha^*-\hat{\boldsymbol{\beta}}_{2\omega}^*\boldsymbol{z}_\alpha^{*(2)})'\right|},\tag{4}$$

其中

$$\begin{aligned}
\boldsymbol{z}_\alpha^{*(2)}&=1,\\
\hat{\boldsymbol{\beta}}_{2\omega}^*&=\hat{\boldsymbol{\nu}}=\bar{\boldsymbol{x}}^*=\bar{\boldsymbol{x}}^{(1)},\\
\hat{\boldsymbol{\beta}}_\Omega^*&=\begin{pmatrix}\hat{\boldsymbol{\beta}}_{1\Omega}^* & \hat{\boldsymbol{\beta}}_{2\Omega}^*\end{pmatrix}\\
&=\left(\sum\boldsymbol{x}_\alpha^*\boldsymbol{z}_\alpha^{*(1)\prime}\sum\boldsymbol{x}_\alpha^*\boldsymbol{z}_\alpha^{*(2)\prime}\right)\begin{pmatrix}\sum\boldsymbol{z}_\alpha^{*(1)}\boldsymbol{z}_\alpha^{*(1)\prime} & \sum\boldsymbol{z}_\alpha^{*(1)}\boldsymbol{z}_\alpha^{*(2)\prime}\\ \sum\boldsymbol{x}_\alpha^{*(2)}\boldsymbol{z}_\alpha^{*(1)\prime} & \sum\boldsymbol{x}_\alpha^{*(2)}\boldsymbol{z}_\alpha^{*(1)\prime}\end{pmatrix}^{-1}\\
&=\begin{pmatrix}\boldsymbol{A}_{12} & N\bar{\boldsymbol{x}}^{(1)}\end{pmatrix}\begin{pmatrix}\boldsymbol{A}_{22} & 0\\ 0 & N\end{pmatrix}^{-1}\\
&=\begin{pmatrix}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} & \bar{\boldsymbol{x}}^{(1)}\end{pmatrix}.
\end{aligned}\tag{5}$$

U 的分母中的矩阵是

$$\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)})'=\boldsymbol{A}_{11}.\tag{6}$$

分子中的矩阵是

$$\begin{aligned}&\sum_{\alpha=1}^N[\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})][\boldsymbol{x}_\alpha^{(1)}-\bar{\boldsymbol{x}}^{(1)}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}(\boldsymbol{x}_\alpha^{(2)}-\bar{\boldsymbol{x}}^{(2)})]'\\
&=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}.\end{aligned}\tag{7}$$

因此

$$U=\frac{|\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|}{|\boldsymbol{A}_{11}|}=\frac{|\boldsymbol{A}|}{|\boldsymbol{A}_{11}|\cdot|\boldsymbol{A}_{22}|},\tag{8}$$

这正是 V.

现在让我们看一下为什么当原假设为真时 $U=V$ 的分布不依赖于 $\boldsymbol{X}_\alpha^{(2)}$ 是否固定. 在第 8 章中已经指出当原假设为真时 U 的分布只依赖于 p,q_1 和 $N-q_2$, 而不依赖于 $\boldsymbol{z}_\alpha$. 因此给定 $\boldsymbol{X}_\alpha^{(2)}=\boldsymbol{x}_\alpha^{(2)}$ 时 V 的条件分布不依赖于 $\boldsymbol{x}_\alpha^{(2)}$; V 和 $\boldsymbol{X}_\alpha^{(2)}$ 的联合分布是 V 的分布和 $\boldsymbol{X}_\alpha^{(2)}$ 的分布的乘积, 并且 V 的边缘分布就是这个条件分布. 这表明 V 的分布 (在原假设下) 和 $\boldsymbol{X}_\alpha^{(2)}$ 是固定的还是服从某个分布 (正态或其他) 没关系.

我们可以把这个结果推广, 当 $q>2$ 时, 在独立性的原假设下 V 的分布不依赖于变量集 $\boldsymbol{X}_{\alpha}^{(1)}$ 的分布. 我们有 $V=V_2\cdots V_q$, 其中 V_i 由 9.3 节 (1) 式定义. 当原假设为真时, 由以前的结果知 V_q 的分布和 $\boldsymbol{X}_{\alpha}^{(1)},\cdots,\boldsymbol{X}_{\alpha}^{(q-1)}$ 独立. 我们依次证明 V_j 的分布和 $\boldsymbol{X}_{\alpha}^{(1)},\cdots,\boldsymbol{X}_{\alpha}^{(j-1)}$ 独立. 因此 $V=V_2\cdots V_q$ 的分布和 $\boldsymbol{X}_{\alpha}^{(1)}$ 独立.

定理 9.8.1 *在独立性的原假设下, 如果 $q-1$ 个集合是联合正态分布的, 则 V 的分布就是本章前面给出的分布, 即使单个集合不服从正态分布.*

在两个变量集的情形时, 我们可能感兴趣的是两个集合之间的关联的一种度量, 它是相关系数的推广. 两个标量 X_1 和 X_2 之间的相关系数的平方可以看成是 X_1 对 X_2 的回归的方差和 X_1 的方差的比, 也就是 $\mathscr{V}(\beta X_2)/\mathscr{V}(X_1)=\beta^2\sigma_{22}/\sigma_{11}=(\sigma_{12}^2/\sigma_{22})/\sigma_{11}=\rho_{12}^2$. 对于向量 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$, 一个相应的度量是 $\boldsymbol{X}^{(1)}$ 对 $\boldsymbol{X}^{(2)}$ 的回归的广义方差和 $\boldsymbol{X}^{(1)}$ 的广义方差的比, 即

$$\frac{|\mathrm{E}[\boldsymbol{\beta}X^{(2)}(\boldsymbol{\beta}X^{(2)})']|}{|\boldsymbol{\Sigma}_{11}|}=\frac{|\boldsymbol{\beta}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}'|}{|\boldsymbol{\Sigma}_{11}|}=\frac{|\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|}{|\boldsymbol{\Sigma}_{11}|}=(-1)^{p_1}\frac{\begin{vmatrix}\mathbf{0}&\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}\end{vmatrix}}{|\boldsymbol{\Sigma}_{11}|}.\tag{9}$$

如果 $p_1=p_2$, 这个度量是

$$\frac{|\boldsymbol{\Sigma}_{12}|^2}{|\boldsymbol{\Sigma}_{11}|\cdot|\boldsymbol{\Sigma}_{22}|}.\tag{10}$$

在某种意义上, 这种度量指出了用 $\boldsymbol{X}^{(2)}$ 预测 $\boldsymbol{X}^{(1)}$ 的好坏程度.

在两个标量的情况下, X_1 和 X_2 的不相关系数是 $\sigma_{1.2}^2/\sigma_1^2$, 其中 $\sigma_{1.2}^2=\mathrm{E}(X_1-\beta X_2)^2$ 是当 $\mathrm{E}(X_1)=\mathrm{E}(X_2)=0$ 且 $\mathrm{E}(X_1|X_2)=\beta X_2$ 时 X_1 对 X_2 的回归的方差. 对于两个向量 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的情况, 回归矩阵是 $\boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$, 且 $\boldsymbol{X}^{(1)}$ 在 $\boldsymbol{X}^{(2)}$ 上的回归的广义方差为

$$|\mathrm{E}\{(\boldsymbol{X}^{(1)}-\boldsymbol{\beta}\boldsymbol{X}^{(2)})(\boldsymbol{X}^{(1)}-\boldsymbol{\beta}\boldsymbol{X}^{(2)})'\}|=|\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|=\frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{22}|}.\tag{11}$$

由于 $\boldsymbol{X}^{(1)}$ 的广义方差是 $|\mathrm{E}(\boldsymbol{X}^{(1)}\boldsymbol{X}^{(1)'})|=|\boldsymbol{\Sigma}_{11}|$, 所以向量不相关系数是

$$\frac{|\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}|}{|\boldsymbol{\Sigma}_{11}|}=\frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{11}|\cdot|\boldsymbol{\Sigma}_{22}|}.\tag{12}$$

和 (12) 式等价的样本不相关系数是 V.

关联的一种度量是 1 减去不相关系数. 这两种度量都可以考虑用分量的个数来作修正. 对于第一种度量, 我们可以取 (9) 式的 p_1 次根; 对于第二种度量, 可以从 1 中减去不相关系数的 p_1 次根. 另一种关联的度量就是

$$\frac{\mathrm{tr}\,\mathrm{E}[\boldsymbol{\beta}\boldsymbol{X}^{(2)}(\boldsymbol{\beta}\boldsymbol{X}^{(2)})'](\mathrm{E}\boldsymbol{X}^{(1)}\boldsymbol{X}^{(1)'})^{-1}}{p}=\frac{\mathrm{tr}\,\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}}{p}.\tag{13}$$

这种度量的范围是在 0 和 1 之间. 如果当 $p_1 \leqslant p_2$ 时 $\boldsymbol{X}^{(1)}$ 可以由 $\boldsymbol{X}^{(2)}$ 精确地预测出来 (即 $\boldsymbol{\Sigma}_{11.2}=\boldsymbol{0}$), 则这个度量就是 1. 如果 $\boldsymbol{X}^{(1)}$ 的线性组合都不能精确地被预测出来, 则度量为 0.

9.9 似然比检验的容许性

在 0-1 损失函数下似然比检验的容许性可以这样证明, 指出它是关于参数的一个合适的先验分布的贝叶斯方法. (见 5.6 节.)

定理 9.9.1 *如果 $N>p+1$, 则 $\boldsymbol{\Sigma}$ 具有 9.2 节中 (6) 式的形式的假设的似然比检验是贝叶斯的和容许的.*

证明 我们将指出似然比检验等价于当

$$\frac{\int f(\boldsymbol{x}|\boldsymbol{\theta})\Pi_1(\mathrm{d}\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta})\Pi_0(\mathrm{d}\boldsymbol{\theta})} \geqslant c \tag{1}$$

时拒绝假设, 其中 $\boldsymbol{x}$ 代表样本, $\boldsymbol{\theta}$ 代表参数 ($\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$), $f(\boldsymbol{x}|\boldsymbol{\theta})$ 是密度, 而 Π_1 和 Π_0 分别正比于备择假设和原假设下 $\boldsymbol{\theta}$ 的概率测度. 具体地说, 不等式左边和 $\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|/|\boldsymbol{A}|$ 的平方根成正比.

为了定义 Π_1, 令

$$\boldsymbol{\mu}=(\boldsymbol{I}+\boldsymbol{V}\boldsymbol{V}')^{-1}\boldsymbol{V}\boldsymbol{Y}, \quad \boldsymbol{\Sigma}=(\boldsymbol{I}+\boldsymbol{V}\boldsymbol{V}')^{-1}, \tag{2}$$

其中 p 元随机向量 $\boldsymbol{V}$ 有正比于 $(1+\boldsymbol{v}'\boldsymbol{v})^{-\frac{1}{2}n}$ 的密度, $n=N-1$, 并且给定 $\boldsymbol{V}=\boldsymbol{v}$ 下 Y 的条件分布是 $N[0,(1+\boldsymbol{v}'\boldsymbol{v})/N]$. 注意到若 $n>q$ 则 $(1+\boldsymbol{v}'\boldsymbol{v})^{-\frac{1}{2}n}$ 的积分是有限的 (见习题 5.15). 则 (1) 式的分子是

$$\begin{aligned}&\mathrm{const}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}'|^{\frac{1}{2}N}\\&\cdot\exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N}\left[\boldsymbol{x}_\alpha-(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v}y\right]'(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')\left[\boldsymbol{x}_\alpha-(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v}y\right]\right\}\\&\cdot(1+\boldsymbol{v}'\boldsymbol{v})^{-\frac{1}{2}n}\cdot(1+\boldsymbol{v}'\boldsymbol{v})^{-\frac{1}{2}}\exp\left\{\frac{-\frac{1}{2}Ny^2}{1+\boldsymbol{v}'\boldsymbol{v}}\right\}\mathrm{d}\boldsymbol{v}\mathrm{d}y.\end{aligned} \tag{3}$$

(3) 式中被积函数的指数是 -2 乘以

$$\begin{aligned}&\sum_{\alpha}^{N}\boldsymbol{x}_\alpha'(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')\boldsymbol{x}_\alpha-2y\boldsymbol{v}'\sum_{\alpha=1}^{N}x_\alpha+Ny^2\boldsymbol{v}'(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v}+\frac{Ny^2}{1+\boldsymbol{v}'\boldsymbol{v}}\\&=\sum_{\alpha=1}^{N}x_\alpha'\boldsymbol{x}_\alpha+\boldsymbol{v}'\sum_{\alpha=1}^{N}x_\alpha'\boldsymbol{x}_\alpha\boldsymbol{v}-2y\boldsymbol{v}'N\bar{\boldsymbol{x}}+Ny^2\\&=\mathrm{tr}\,\boldsymbol{A}+\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}+N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}+N(y-\bar{\boldsymbol{x}}'\boldsymbol{v})^2,\end{aligned} \tag{4}$$

其中 $\boldsymbol{A}=\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha\boldsymbol{x}_\alpha'-N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'$. 我们用到了 $\boldsymbol{v}'(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')^{-1}\boldsymbol{v}+(1+\boldsymbol{v}'\boldsymbol{v})^{-1}=1$ [由 $(\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}')^{-1}=\boldsymbol{I}-(1+\boldsymbol{v}'\boldsymbol{v})^{-1}\boldsymbol{v}\boldsymbol{v}'$ 得到]. 利用 $|\boldsymbol{I}+\boldsymbol{v}\boldsymbol{v}'|=1+\boldsymbol{v}'\boldsymbol{v}$ (推论 A.3.1), 我们写 (3) 式为

$$\text{const}\,\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{A}-\frac{1}{2}N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}}\mathrm{d}\boldsymbol{v}=\text{const}\,|\boldsymbol{A}|^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{A}-\frac{1}{2}N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}. \tag{5}$$

为了定义 Π_0, 令 $\boldsymbol{\Sigma}$ 有 9.2 节中 (6) 式的形式. 令

$$\left[\boldsymbol{\mu}^{(i)},\boldsymbol{\Sigma}_{ii}\right]=\left[(\boldsymbol{I}+\boldsymbol{V}^{(i)}\boldsymbol{V}^{(i)\prime})^{-1}\boldsymbol{V}^{(i)}\boldsymbol{Y}_i,(\boldsymbol{I}+\boldsymbol{V}^{(i)}\boldsymbol{V}^{(i)\prime})^{-1}\right],\quad i=1,\cdots,q, \tag{6}$$

其中 p_i 元随机向量 $\boldsymbol{V}^{(i)}$ 有正比于 $(1+\boldsymbol{v}^{(i)\prime}\boldsymbol{v}^{(i)})^{-\frac{1}{2}n}$ 的密度, 并且给定 $\boldsymbol{V}^{(i)}=\boldsymbol{v}^{(i)}$ 下 Y_i 的条件分布是 $N[0,(1+\boldsymbol{v}^{(i)\prime}\boldsymbol{v}^{(i)})/N]$, 令 $(\boldsymbol{V}_1,Y_1),\cdots,(\boldsymbol{V}_q,Y_q)$ 相互独立. 则 (1) 式的分母是

$$\begin{aligned}&\prod_{i=1}^{q}\text{const}\,|\boldsymbol{A}_{ii}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathrm{tr}\,\boldsymbol{A}_{ii}+N\bar{\boldsymbol{x}}^{(i)\prime}\bar{\boldsymbol{x}}^{(i)})\right] &(7)\\ &=\text{const}\left(\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|^{-\frac{1}{2}}\right)\exp\left[-\frac{1}{2}(\mathrm{tr}\,\boldsymbol{A}+N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}})\right].\end{aligned}$$

则 (1) 式的左边正比于 $\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|/|\boldsymbol{A}|$ 的平方根. ■

这个证明来自 Kiefer and Schwartz(1965).

9.10　子集间独立性检验的功效函数的单调性

令 $\boldsymbol{Z}_\alpha=[\boldsymbol{Z}_\alpha^{(1)\prime},\boldsymbol{Z}_\alpha^{(2)\prime}]'\,(\alpha=1,\cdots,n)$ 服从分布

$$N\left[\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{11}&\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}\end{pmatrix}\right]. \tag{1}$$

我们想检验 $H:\boldsymbol{\Sigma}_{12}=\boldsymbol{0}$. 不失一般性我们假设 $p_1\leqslant p_2$. 令 $\rho_1,\cdots,\rho_{p_1}(\rho_1\geqslant\cdots\geqslant\rho_{p_1})$ 为 (总体) 典型相关系数. (这些 ρ_i^2 是 $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 的特征根, 见第 12 章.) 令 $\boldsymbol{R}=\mathrm{diag}(\rho_1,\cdots,\rho_{p_1})$, $\boldsymbol{\Delta}=[\boldsymbol{R},\boldsymbol{0}](p_1\times p_2)$.

引理 9.10.1　*存在矩阵* $\boldsymbol{B}_1\,(p_1\times p_1)$, $\boldsymbol{B}_2\,(p_2\times p_2)$ *使得*

$$\boldsymbol{B}_1\boldsymbol{\Sigma}_{11}\boldsymbol{B}_1'=\boldsymbol{I}_{p_1},\quad \boldsymbol{B}_2\boldsymbol{\Sigma}_{22}\boldsymbol{B}_2'=\boldsymbol{I}_{p_2},\quad \boldsymbol{B}_1\boldsymbol{\Sigma}_{12}\boldsymbol{B}_2'=\boldsymbol{\Delta}. \tag{2}$$

证明　在引理 8.10.13 中令 $m=p_2$, $\boldsymbol{B}=\boldsymbol{B}_1$, $\boldsymbol{F}'=\boldsymbol{\Sigma}_{22}^{\frac{1}{2}}\boldsymbol{B}_2'$, $\Xi=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}}$. 则 $\boldsymbol{F}'\boldsymbol{F}=\boldsymbol{B}_2\boldsymbol{\Sigma}_{22}\boldsymbol{B}_2'=\boldsymbol{I}_{p_2}$, $\boldsymbol{B}_1\boldsymbol{\Sigma}_{12}\boldsymbol{B}_2'=\boldsymbol{B}_1\Xi\boldsymbol{F}=\boldsymbol{\Delta}$. ■

(这个引理也包含在 12.2 节中.)

令 $\boldsymbol{x}_\alpha=\boldsymbol{B}_1\boldsymbol{Z}_\alpha^{(1)}$, $y_\alpha=\boldsymbol{B}_2\boldsymbol{Z}_\alpha^{(2)}$, $\alpha=1,\cdots,n$, 且 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$, $\boldsymbol{Y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)$. 则 $(\boldsymbol{x}_\alpha',\boldsymbol{y}_\alpha')'\,(\alpha=1,\cdots,n)$ 独立并且服从分布

$$N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{\Delta} \\ \boldsymbol{\Delta}' & \boldsymbol{I}_q \end{pmatrix}\right]. \tag{3}$$

假设 $H:\boldsymbol{\Sigma}_{12}=0$ 等价于 $H:\boldsymbol{\Delta}=\mathbf{0}$(即所有的典型相关系数 $\rho_1,\cdots,\rho_{p_1}$ 都是 0). 现在给定 $\boldsymbol{Y}$, 向量 $\boldsymbol{x}_\alpha\,(\alpha=1,\cdots,n)$ 条件独立并且服从分布 $N(\boldsymbol{\Delta y}_\alpha,\boldsymbol{I}-\boldsymbol{\Delta\Delta}')=N(\boldsymbol{\Delta y}_\alpha,\boldsymbol{I}-\boldsymbol{R}^2)$. 则 $\boldsymbol{x}_\alpha^*=(\boldsymbol{I}_{p_1}-\boldsymbol{R}^2)^{-\frac{1}{2}}\boldsymbol{x}_\alpha$ 服从 $N(\boldsymbol{My}_\alpha,\boldsymbol{I}_{p_1})$, 其中

$$\begin{aligned} \boldsymbol{M}&=(\boldsymbol{D},\mathbf{0}),\\ \boldsymbol{D}&=\operatorname{diag}(\delta_1,\cdots,\delta_{p_1}),\\ \delta_i&=\rho_i/(1-\rho_i^2)^{\frac{1}{2}},\quad i=1,\cdots,p_1. \end{aligned} \tag{4}$$

需要指出 δ_i^2 是 $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11.2}^{-1}$ 的特征根, 其中 $\boldsymbol{\Sigma}_{11.2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

不变检验只依赖于 (样本) 典型相关系数 $r_i=\sqrt{c_i}$, 其中

$$c_i=\lambda_i\left[(\boldsymbol{X}^*\boldsymbol{X}^{*'})^{-1}(\boldsymbol{X}^*\boldsymbol{Y}')(\boldsymbol{YY}')^{-1}(\boldsymbol{YX}^{*'})\right]. \tag{5}$$

令

$$\begin{aligned} \boldsymbol{S}_h&=\boldsymbol{X}^*\boldsymbol{Y}'(\boldsymbol{YY}')^{-1}\boldsymbol{YX}^{*'},\\ \boldsymbol{S}_e&=\boldsymbol{X}^*\boldsymbol{X}^{*'}-\boldsymbol{S}_h=\boldsymbol{X}^*\left[\boldsymbol{I}-\boldsymbol{Y}'(\boldsymbol{YY}')^{-1}\boldsymbol{Y}\right]\boldsymbol{X}^{*'}. \end{aligned} \tag{6}$$

则

$$\lambda_i(\boldsymbol{S}_h\boldsymbol{S}_e^{-1})=\frac{c}{1-c_i}. \tag{7}$$

现在给定 $\boldsymbol{Y}$, 问题就简化为 MANOVA 问题, 并且我们可以如下应用定理 8.10.6. 存在一个正交变换 (见 8.3.3 节) 把 $\boldsymbol{X}_*$ 变为 $(\boldsymbol{U},\boldsymbol{V})$, 从而 $\boldsymbol{S}_h=\boldsymbol{UU}'$, $\boldsymbol{S}_e=\boldsymbol{VV}'$, $\boldsymbol{U}=(\boldsymbol{U}_1,\cdots,\boldsymbol{u}_{p_2})$, $\boldsymbol{V}$ 是 $p_1\times(n-p_2)$ 维的, 当 $i=1,\cdots,p_1$ 时, $\boldsymbol{u}_i$ 服从正态分布 $N(\delta_i\boldsymbol{\varepsilon}_i,\boldsymbol{I})$ ($\boldsymbol{\varepsilon}_i$ 是 $\boldsymbol{I}$ 的第 i 列), 当 $i=p_1+1,\cdots,p_2$ 时, 它服从正态分布 $N(0,\boldsymbol{I})$, 并且 $\boldsymbol{V}$ 的列独立且都服从分布 $N(0,\boldsymbol{I})$. 则 $c_1,\cdots,c_{p_1}$ 是 $\boldsymbol{UU}'(\boldsymbol{VV}')^{-1}$ 的特征根, 并且它们的分布依赖于 $\boldsymbol{MYY}'\boldsymbol{M}'$ 的特征根 $\tau_1^2,\cdots,\tau_{p_1}^2$. 现在从定理 8.10.6, 我们得到下面的引理.

引理 9.10.2 *在给定 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 的其他列的条件下, 如果一个不变检验的接受区域对于 $\boldsymbol{U}$ 的每一列都是凸的, 则给定 $\boldsymbol{Y}$ 下其条件功效对于 $\boldsymbol{MYY}'\boldsymbol{M}'$ 的每一个特征根 τ_i^2 是增加的.*

引理 9.10.3 *如果 $\boldsymbol{A}\geqslant\boldsymbol{B}$, 则 $\lambda_i(\boldsymbol{A})\geqslant\lambda_i(\boldsymbol{B})$.*

证明 由特征根的极小化极大性质 [例如参见 Courant and Hilbert(1953)],

$$\lambda_i(\boldsymbol{A})=\max_{S_i}\min_{\boldsymbol{x}\in S_i}\frac{\boldsymbol{x}'\boldsymbol{Ax}}{\boldsymbol{x}'\boldsymbol{x}}\geqslant\max_{S_i}\min_{\boldsymbol{x}\in S_i}\frac{\boldsymbol{x}'\boldsymbol{Bx}}{\boldsymbol{x}'\boldsymbol{x}}=\lambda_i(\boldsymbol{B}), \tag{8}$$

其中 $\boldsymbol{S}_i$ 取值于 i 维子空间. ■

现在把引理 9.10.3 应用于 $\boldsymbol{MYY'M'}$, 可以看出, 对于每个 j, τ_i^2 是 $\delta_i = \rho_i/(1-\rho_i^2)^{\frac{1}{2}}$ 的增函数, 从而也是 ρ_i 的增函数. 由于 $\boldsymbol{Y}$ 的边缘分布不依赖于 ρ_i, 取其无条件功效我们得到下面的定理.

定理 9.10.1　在给定 $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 的其他列的条件下, 如果一个不变检验的接受区域对于 $\boldsymbol{U}$ 的每一列都是凸的, 则其功效函数对于每个 ρ_i 是单增的.

9.11　椭球等高分布

9.11.1　椭球等高分布的观测

令 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是随机向量 $\boldsymbol{X}$ 的 N 个观测, 且 $\boldsymbol{X}$ 的密度为

$$|\boldsymbol{\Lambda}|^{-\frac{1}{2}}g\left[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})\right], \tag{1}$$

其中 $\mathrm{E}(R^4)<\infty$, 且 $R^2=(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})$. 则 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{\nu}$ 且 $\mathrm{E}(\boldsymbol{X}-\boldsymbol{\nu})(\boldsymbol{X}-\boldsymbol{\nu})'=\boldsymbol{\Sigma}=(\mathrm{E}(R^2)/p)\boldsymbol{\Lambda}$. 令

$$\bar{\boldsymbol{x}}=\frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha,\quad \boldsymbol{S}=\frac{1}{N-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})'. \tag{2}$$

则

$$\sqrt{N}\mathrm{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\xrightarrow{d}N\left[\boldsymbol{0},(\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\kappa\,\mathrm{vec}\,\boldsymbol{\Sigma}(\mathrm{vec}\,\boldsymbol{\Sigma})'\right], \tag{3}$$

其中 $1+\kappa=p\mathrm{E}(R^4)/[(p+2)(\mathrm{E}(R^2))^2]$.

检验原假设 $\boldsymbol{\Sigma}_{ij}=\boldsymbol{0}\,(i\neq j)$ 的似然比准则是 $U=\prod_{i=2}^{q}V_i$ 的 $N/2$ 次幂, 其中 V_i 是检验原假设 $\boldsymbol{\Sigma}_{1i}=\boldsymbol{0},\cdots,\boldsymbol{\Sigma}_{i-1,i}=\boldsymbol{0}$ 的 U 准则并且由 9.3 节的 (1) 式和 (6) 式给出. V_i 的形式是第 8 章中的似然比准则 U, 只是把 $\boldsymbol{X}$ 用 $\boldsymbol{X}^{(i)}$ 来代替, $\boldsymbol{\mathsf{B}}$ 用 9.3 节的 (5) 式给出的 $\boldsymbol{\mathsf{B}}_i$ 代替, 在原假设 $\boldsymbol{\mathsf{B}}_i=\boldsymbol{0}$ 下 $\boldsymbol{Z}$ 用

$$\tilde{\boldsymbol{X}}^{(i-1)}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \vdots\\ \boldsymbol{X}^{(i-1)}\end{bmatrix} \tag{4}$$

来代替, $\boldsymbol{\Sigma}$ 用 $\boldsymbol{\Sigma}_{ii}$ 来代替. 子向量 $\tilde{\boldsymbol{X}}^{(i-1)}$ 和 $\boldsymbol{X}^{(i)}$ 不相关, 但是和 $\boldsymbol{X}^{(i)}$ 不独立, 除非 $(\tilde{\boldsymbol{X}}^{(i-1)'},\boldsymbol{X}^{(i)'})'$ 是正态的. 令

$$\tilde{\boldsymbol{A}}^{(i-1)}=\begin{bmatrix}\boldsymbol{A}_{11}&\cdots&\boldsymbol{A}_{1,i-1}\\ \vdots&&\vdots\\ \boldsymbol{A}_{i-1,1}&\cdots&\boldsymbol{A}_{i-1,i-1}\end{bmatrix}, \tag{5}$$

$$\tilde{\boldsymbol{A}}^{(i,i-1)}=(\boldsymbol{A}_{i1},\cdots,\boldsymbol{A}_{i,i-1})=\tilde{\boldsymbol{A}}^{(i-1,i)'}, \tag{6}$$

$\tilde{\boldsymbol{\Sigma}}^{(i-1)}, \tilde{\boldsymbol{\Sigma}}^{(i,i-1)}, \tilde{\boldsymbol{S}}^{(i-1)}$ 和 $\tilde{\boldsymbol{S}}^{(i,i-1)}$ 也有类似的定义. 我们记 $V_i = |\boldsymbol{G}_i|/|\boldsymbol{G}_i + \boldsymbol{H}_i|$, 其中

$$\boldsymbol{H}_i = \tilde{\boldsymbol{A}}^{(i,i-1)}(\tilde{\boldsymbol{A}}^{(i-1)})^{-1}\tilde{\boldsymbol{A}}^{(i-1,i)} \tag{7}$$

$$= (N-1)\tilde{\boldsymbol{S}}^{(i,i-1)}(\tilde{\boldsymbol{S}}^{(i-1)})^{-1}\tilde{\boldsymbol{S}}^{(i-1,i)},$$

$$\boldsymbol{G}_i = \boldsymbol{A}_{ii} - \boldsymbol{H}_i = (N-1)\boldsymbol{S}_{ii} - \boldsymbol{H}_i. \tag{8}$$

定理 9.11.1 *当 $\boldsymbol{X}$ 有密度 (1) 并且原假设为真时, $\boldsymbol{H}_i$ 的极限分布是 $W[(1+\kappa)\boldsymbol{\Sigma}_{ii}, \bar{p}_i]$, 其中 $\bar{p}_i = p_1 + \cdots + p_{i-1}$ 且 p_j 是 $\boldsymbol{X}^{(j)}$ 的分量数.*

证明 由于 $\tilde{\boldsymbol{\Sigma}}^{(i,i-1)} = \boldsymbol{0}$, 我们有 $\mathrm{E}(\tilde{\boldsymbol{S}}^{(i,i-1)}) = \boldsymbol{0}$, 并且如果 $j,l \leqslant \bar{p}_i$ 且 $k,m > \bar{p}_i$ 或者 $j,l > \bar{p}_i$ 且 $k,m \leqslant \bar{p}_i$, 则

$$\mathrm{E}(s_{jk}s_{lm}) = \left(\frac{\kappa}{N} + \frac{1}{N-1}\right)\sigma_{jl}\sigma_{km}, \tag{9}$$

否则 $\mathrm{E}(s_{jk}s_{lm}) = 0$(定理 3.6.1). 我们可以记

$$\mathrm{E}[\mathrm{vec}\,\tilde{\boldsymbol{S}}^{(i,i-1)}(\mathrm{vec}\,\tilde{\boldsymbol{S}}^{(i,i-1)})'] = \left(\frac{\kappa}{N} + \frac{1}{N-1}\right)(\boldsymbol{\Sigma}_{ii} \otimes \tilde{\boldsymbol{\Sigma}}^{(i-1)}). \tag{10}$$

由于 $\boldsymbol{S}^{(i-1)} \to_p \boldsymbol{\Sigma}^{(i-1)}$ 和 $\sqrt{N}\mathrm{vec}\,\boldsymbol{S}^{(i,i-1)}$ 的极限分布是正态, 定理 9.10.1 可以由 8.4 节的 (2) 式得到. ■

定理 9.11.2 *在定理 9.11.1 的条件下, 当原假设为真时,*

$$-N\ln V_i \xrightarrow{d} (1+\kappa)\chi^2_{\bar{p}_i p_i}. \tag{11}$$

证明 我们可以记 $V_i = |\boldsymbol{I} + N^{-1}(\frac{1}{N}\boldsymbol{G}_i)^{-1}\boldsymbol{H}_i|$ 且利用 $N\ln|\boldsymbol{I} + N^{-1}\boldsymbol{C}| = \mathrm{tr}\,\boldsymbol{C} + O_p(N^{-1})$ 和

$$\mathrm{tr}\left(\frac{1}{N}\boldsymbol{G}_i\right)^{-1}\boldsymbol{H}_i = N\sum_{i,j=\bar{p}_i+1}^{\bar{p}_{i+1}}\sum_{g,h=1}^{\bar{p}_i} g^{ij}s_{ig}\bar{s}^{gh}s_{jh} \tag{12}$$

$$= N(\mathrm{vec}\,\boldsymbol{S}^{(i,i-1)})'\left[\left(\frac{1}{N}\boldsymbol{G}_i\right)^{-1} \otimes \tilde{\boldsymbol{S}}_{ii}^{-1}\right]\mathrm{vec}\,\boldsymbol{S}^{(i-1,i)}.$$ ■

因为在原假设 $H_i : \tilde{\boldsymbol{\Sigma}}^{(i,i-1)} = \boldsymbol{0}$ 为真时, $\boldsymbol{X}^{(i)}$ 和 $\tilde{\boldsymbol{X}}^{(i-1)}$ 不相关, 所以 V_i 和 $V_2, \cdots, V_{i-1}$ 渐近独立. 在原假设 $H_2, \cdots, H_i$ 为真时, V_i 和 $V_2, \cdots, V_{i-1}$ 渐近独立. 由定理 9.10.2 得到

$$-N\ln V = -N\sum_{i=2}^{q}\ln V_i \xrightarrow{d} \chi^2_f, \tag{13}$$

其中 $f = \sum_{i=2}^{p}\bar{p}_i p_i = \frac{1}{2}[p(p+1) - \sum_{i=1}^{q}p_i(p_i+1)]$. 9.2 节的似然比检验可以在渐近的条件下实现.

令 $A_0 = \mathrm{diag}(A_{11}, \cdots, A_{qq})$. 则当 $\Sigma = \mathrm{diag}(\Sigma_{11}, \cdots, \Sigma_{qq})$ 时

$$\frac{1}{2}\mathrm{tr}\,(AA_0^{-1} - I)^2 = \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{g} \mathrm{tr}\, A_{ij}A_{jj}^{-1}A_{ji}A_{ii}^{-1} \tag{14}$$

有 χ_f^2 – 分布. 9.6.1 节的逐步下降法在渐近的条件下也是正确的.

9.11.2 椭球等高矩阵分布

令 $y(p \times N)$ 有密度 $g(\mathrm{tr}\, YY')$. 矩阵 Y 是向量–球形的, 也就是说, $\mathrm{vec}\, Y$ 是球形的且有随机表示 $\mathrm{vec}\, Y = R\,\mathrm{vec}\, U_{p\times N}$, 其中 $R^2 = (\mathrm{vec}\, Y)'\mathrm{vec}\, Y = \mathrm{tr}\, YY'$ 且 $\mathrm{vec}\, U_{p\times N}$ 在单位球面 $(\mathrm{vec}\, U_{p\times N})'\mathrm{vec}\, U_{p\times N} = 1$ 上均匀分布. (我们利用符号 $U_{p\times N}$ 来区别空间 $UU' = I_p$ 上的 U 均匀分布.)

令

$$X = \nu\varepsilon_N' + CY, \tag{15}$$

其中 $\Lambda = CC'$ 且 C 是下三角的. 则 X 有密度

$$|\Lambda|^{-N/2} g\left[\mathrm{tr}\, C^{-1}(X - \nu\varepsilon_N')(X' - \varepsilon_N\nu')(C')\right]^{-1} \tag{16}$$

$$= |\Lambda|^{-N/2} g\left[\mathrm{tr}\,(X' - \varepsilon_N\nu')\Lambda^{-1}(X - \nu\varepsilon_N')\right]. \tag{17}$$

考虑原假设 $\Sigma_{ij} = 0$, $i \neq j$, 或者等价地 $\Lambda_{ij} = 0$, $i \neq j$, 或者等价地 $R_{ij} = 0$, $i \neq j$. 则 $C = \mathrm{diag}\,(C_{11}, \cdots, C_{qq})$.

令 $M = I_N - (1/N)\varepsilon_N\varepsilon_N'$, 由于 $M^2 = M$, M 是一个有 $N-1$ 个特征根 1 和一个特征根 0 的幂等阵. 则 $A = XMX'$ 且 $A_{ii} = A^{(i)}MX^{(i)\prime}$. 似然函数是

$$|\Lambda|^{-n/2} g\{\mathrm{tr}\,\Lambda^{-1}[A + N(\bar{x} - \nu)(\bar{x} - \nu)']\}. \tag{18}$$

矩阵 A 和向量 $\bar{x}$ 是充分统计量, 假设 H 的似然比准则是 $(|A|/\prod_{i=1}^{g}|A_{ii}|)^{N/2}$, 对于正态的情况也一样. 参见 Anderson and Fang(1990b).

定理 9.11.3 *令 $f(X)$ 是 $X\,(p \times N)$ 的向量值函数, 使得对于所有的 ν 有*

$$f(X + \nu\varepsilon_N') = f(X), \tag{19}$$

并且对于所有的 $K = \mathrm{diag}\,(K_{11}, \cdots, K_{qq})$ 有

$$f(KX) = f(X). \tag{20}$$

则当 X 有任意密度 (16) 式时 $f(X)$ 的分布和当 X 有正态密度 (16) 时 $f(X)$ 的分布相同.

证明 证明类似于定理 4.5.4 的证明. ■

由定理 9.11.3 可以看出, 由于 V 在变换 $X \to KX$ 下是不变的, 所以在原假设 H 下当 X 有密度 (16) 式时 V 的分布和当 X 有正态分布时是相等的. 类似地, 由于 V_1 和准则 (14) 是不变的, 所以都有正态分布.

习　题

9.1　(9.3 节)　利用 $V^h w(\boldsymbol{A}|\boldsymbol{\Sigma}_0, n)$ 的积分证明

$$\mathrm{E}(V^h) = \frac{\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(n+1-i)+h\right]\prod_{i=1}^{q}\{\prod_{j=1}^{p_i}\Gamma\left[\frac{1}{2}(n+1-j)\right]\}}{\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(n+1-i)\right]\prod_{i=1}^{q}\{\prod_{j=1}^{p_i}\Gamma\left[\frac{1}{2}(n+1-j)+h\right]\}}.$$

提示: 证明

$$\mathrm{E}(V^h) = \frac{K(\boldsymbol{\Sigma}_0, n)}{K(\boldsymbol{\Sigma}_0, n+2h)}\int\cdots\int\prod_{i=1}^{q}|\boldsymbol{A}_{ii}|^{-h} w(\boldsymbol{A}, \boldsymbol{\Sigma}_0, n+2h)\mathrm{d}\boldsymbol{A},$$

其中 $K(\boldsymbol{\Sigma}, n)$ 由 $w(\boldsymbol{A}|\boldsymbol{\Sigma}, n) = K(\boldsymbol{\Sigma}, n)|\boldsymbol{A}|^{-\frac{1}{2}(n-p-1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{A}}$ 定义. 利用定理 7.3.5 来证明

$$\mathrm{E}(V^h) = \frac{K(\boldsymbol{\Sigma}_0, n)}{K(\boldsymbol{\Sigma}_0, n+2h)}\prod_{i=1}^{q}\left[\frac{K(\boldsymbol{\Sigma}_{ii}, n+2h)}{K(\boldsymbol{\Sigma}_{ii}, n)}\int\cdots\int w(\boldsymbol{A}_{ii}|\boldsymbol{\Sigma}_{ii}, n)\mathrm{d}\boldsymbol{A}_{ii}\right].$$

9.2　(9.3 节)　证明: 如果 $p_1 = p_2 = p_3 = 1$ [Wilks(1935)], 则

$$\Pr\{V \leqslant v\} = I_v\left[\frac{1}{2}(n-1), \frac{1}{2}\right] + 2B^{-1}\left[\frac{1}{2}(n-1), \frac{1}{2}\right]\sin^{-1}\sqrt{1-v}.$$

[**提示**: 利用定理 9.3.3 和 $\Pr\{V \leqslant \nu\} = 1 - \Pr\{v \leqslant V\}$.]

9.3　(9.3 节)　证明: 如果 $p_1 = p_2 = p_3 = 2$ [Wilks(1935)], 则

$$\begin{aligned}\Pr\{V \leqslant v\} = {} & I_{\sqrt{v}}(n-5, 4) \\ & + B^{-1}(n-5, 4)v^{\frac{1}{2}(n-5)}\left\{n/6 - \frac{1}{2}(n-1)\sqrt{v} - \frac{3}{2}(n-4)v\right. \\ & \left. + \left(\frac{17}{6}n - \frac{15}{2}\right)v^{3/2} - \frac{3}{2}(n-2)v\ln v - \frac{1}{2}(n-3)v^{3/2}\ln v\right\}.\end{aligned}$$

[**提示**: 利用 (9) 式.]

9.4　(9.3 节)　自己推导由 Wilks(1935) 得到和 9.3.3 节最后提到的的一些分布. [**提示**: 除了利用习题 9.2 和习题 9.3 的结论外, 也利用 9.3.2 节的一些结果.]

9.5　(9.4 节)　对 $p_i = 2$ 的情况, 写出 k 和 γ_2 的表达式. 对于 (6) 式, 在 $p = 4$ 和 6 且 $N = 15$ 时, 并且当 ν 的选择使得第一项是 0.95 时, 计算第二项.

9.6　(9.5 节)　证明: 对于正定矩阵 $\boldsymbol{A}$ 和非奇异矩阵 $\boldsymbol{B}$ 和 $\boldsymbol{C}$, 如果 $\boldsymbol{BAB}' = \boldsymbol{CAC}' = \boldsymbol{I}$, 则 $\boldsymbol{B} = \boldsymbol{QC}$, 其中 $\boldsymbol{Q}$ 是正交矩阵.

9.7　(9.5 节)　证明在原假设下 N 乘以 (2) 式的极限分布是 χ^2 分布, 且自由度为 f.

9.8　(9.8 节)　给出样本的向量不相关系数和向量相关系数.

9.9　(9.8 节)　如果 y 是样本向量不相关系数且 z 是向量相关系数的平方, 当 $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$ 时, 求出 $\mathrm{E}(y^g z^h)$.

9.10　(9.9 节)　证明: 如果 $p < n$, 则

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(1+\sum_{i=1}^{p}v_i^2)^{\frac{1}{2}r}}\mathrm{d}v_1\cdots\mathrm{d}v_p < \infty.$$

[**提示**: 依次令 $y_j = w_j\sqrt{1+\sum_{i=j+1}^{p}y_i^2}$, $j = 1, \cdots, p-1$.]

9.11 设 x_1 = 计算速度, x_2 = 计算能力, x_3 = 智力因素, x_4 = 社会因素, x_5 = 活动因素. 基于对 109 个学生的一系列测试, Kelley(1928) 得到如下的相关矩阵:

$$\begin{pmatrix} 1.0000 & 0.4249 & -0.0552 & -0.0031 & 0.1927 \\ 0.4249 & 1.0000 & -0.0416 & 0.0495 & 0.0687 \\ -0.0552 & -0.0416 & 1.0000 & 0.7474 & 0.1691 \\ -0.0031 & 0.0495 & 0.7474 & 1.0000 & 0.2653 \\ 0.1927 & 0.0687 & 0.1691 & 0.2653 & 1.0000 \end{pmatrix}.$$

令 $\boldsymbol{x}^{(1)'} = (x_1, x_2), \boldsymbol{x}^{(2)'} = (x_3, x_4, x_5)$. 在显著性水平 1% 下检验假设 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 独立.

9.12 对习题 3.42 进行相同的练习.

9.13 基于 108 个观测的另一个时间研究数据 [Abruzzi(1950)] 以相关阵形式总结如下:

$$\begin{pmatrix} 1.00 & -0.27 & 0.06 & 0.07 & 0.02 \\ -0.27 & 1.00 & -0.01 & -0.02 & -0.02 \\ 0.06 & -0.01 & 1.00 & -0.07 & -0.04 \\ 0.07 & -0.02 & -0.07 & 1.00 & -0.10 \\ 0.02 & -0.02 & -0.04 & -0.10 & 1.00 \end{pmatrix}.$$

在显著性水平 5% 下检验假设 $\sigma_{ij} = 0, i \neq j$.

第10章　协方差阵相等以及均值向量和协方差阵均相等的假设检验

10.1　引　言

在本章中我们研究协方差阵相等及协方差阵和均值向量都相等的假设检验问题. 在每种情况下 (除了一元), 考虑的问题和检验都是一元问题和检验的多元推广. 很多检验都是似然比检验或者调整后的似然比检验. 不变性的考虑可以引出其他检验过程.

首先, 我们考虑在没有确定公共协方差阵或公共协方差阵和均值向量的情况下, 几个总体的协方差阵相等及协方差阵和均值向量相等的问题. 本章考虑随机因子的多元方差分析问题. 然后我们处理协方差阵与给定矩阵相等以及协方差阵与给定矩阵相等同时均值向量与给定的向量相等的问题. 考虑的另一个假设, 协方差阵和给定的矩阵仅差一个比例系数的情况, 只有一个平凡的相应单变量假设.

在每种情况下, 一个假设类的检验类可以产生一个置信区域. 本章还将给出协方差和协方差的比例的联合置信区间类.

椭球等高分布检验的应用将在 10.11 节给出.

10.2　检验几个协方差阵相等的准则

在本节中, 我们研究几个正态分布, 并且考虑利用从每个总体中得到的一个样本集来检验这些总体的协方差阵相等的假设. 令 $\boldsymbol{x}_\alpha^{(g)}$ $(\alpha=1,\cdots,N_g,\ g=1,\cdots,q)$ 为从第 g 个总体 $N(\boldsymbol{\mu}^{(g)},\boldsymbol{\Sigma}_g)$ 中得到的一个观测. 我们希望检验假设

$$H_1:\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_q. \tag{1}$$

令 $\sum_{g=1}^q N_g=N$,

$$\begin{aligned}\boldsymbol{A}_g&=\sum_{\alpha=1}^{N_g}(\boldsymbol{x}_\alpha^{(g)}-\bar{\boldsymbol{x}}^{(g)})(\boldsymbol{x}_\alpha^{(g)}-\bar{\boldsymbol{x}}^{(g)})',\quad g=1,\cdots,q,\\ \boldsymbol{A}&=\sum_{g=1}^{q}\boldsymbol{A}_g.\end{aligned} \tag{2}$$

首先我们可以得到似然比准则. 似然函数是

$$L=\prod_{g=1}^{q}\frac{1}{(2\pi)^{\frac{1}{2}pN_g}|\boldsymbol{\Sigma}_g|^{\frac{1}{2}N_g}}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N_g}(\boldsymbol{x}_\alpha^{(g)}-\boldsymbol{\mu}^{(g)})'\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{x}_\alpha^{(g)}-\boldsymbol{\mu}^{(g)})\right]. \tag{3}$$

空间 Ω 是使 $\boldsymbol{\Sigma}_g$ 为正定的且 $\boldsymbol{\mu}^{(g)}$ 为任意向量的参数空间. 空间 ω 是使 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\cdots=\boldsymbol{\Sigma}_q$(正定) 且 $\boldsymbol{\mu}^{(g)}$ 为任意向量的参数空间. Ω 中 $\boldsymbol{\mu}^{(g)}$ 和 $\boldsymbol{\Sigma}_g$ 的极大似然估计由下式给出,

$$\hat{\boldsymbol{\mu}}_\Omega^{(g)}=\bar{\boldsymbol{x}}^{(g)},\quad \hat{\boldsymbol{\Sigma}}_{g\Omega}=\frac{1}{N_g}\boldsymbol{A}_g,\quad g=1,\cdots,q. \tag{4}$$

由于不管 $\boldsymbol{\Sigma}_g$ 取什么值 $\boldsymbol{\mu}^{(g)}$ 的极大值都是相同的, 所以在 ω 中 $\boldsymbol{\mu}^{(g)}$ 的极大似然估计也由 (4) 式给出, $\hat{\boldsymbol{\mu}}_\omega^{(g)}=\bar{\boldsymbol{x}}^{(g)}$. 关于 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\cdots=\boldsymbol{\Sigma}_q=\boldsymbol{\Sigma}$ 的极大化方程是

$$\frac{1}{(2\pi)^{\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{\frac{1}{2}N}}\exp\left[-\frac{1}{2}\sum_{g=1}^{q}\sum_{\alpha=1}^{N_g}(\boldsymbol{x}_\alpha^{(g)}-\bar{\boldsymbol{x}}^{(g)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha^{(g)}-\bar{\boldsymbol{x}}^{(g)})\right]. \tag{5}$$

由引理 3.2.2, $\boldsymbol{\Sigma}$ 的极大值是

$$\hat{\boldsymbol{\Sigma}}_\omega=\frac{1}{N}\boldsymbol{A}, \tag{6}$$

且似然函数的极大值为

$$\frac{1}{(2\pi)^{\frac{1}{2}pN}|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}\mathrm{e}^{-\frac{1}{2}pN}. \tag{7}$$

检验 (1) 的似然比准则是

$$\lambda_1=\frac{\prod_{g=1}^{q}|\hat{\boldsymbol{\Sigma}}_{g\Omega}|^{\frac{1}{2}N_g}}{|\hat{\boldsymbol{\Sigma}}_\omega|^{\frac{1}{2}N}}=\frac{\prod_{g=1}^{q}|\boldsymbol{A}_g|^{\frac{1}{2}N_g}}{|\boldsymbol{A}|^{\frac{1}{2}N}}\cdot\frac{N^{\frac{1}{2}pN}}{\prod_{g=1}^{q}N_g^{\frac{1}{2}pN_g}}. \tag{8}$$

临界区域是

$$\lambda_1\leqslant\lambda_1(\varepsilon), \tag{9}$$

其中 $\lambda_1(\varepsilon)$ 的定义是为了使当 (1) 为真时 (9) 式以概率 ε 成立.

Bartlett(1973a) 建议过在单变量情形下调整 λ_1, 用 A_g 的自由度的数目来代替样本数. 在差一个常数倍的情形下, 他提出的统计量是

$$V_1=\frac{\prod_{g=1}^{q}|\boldsymbol{A}_g|^{\frac{1}{2}n_g}}{|\boldsymbol{A}|^{\frac{1}{2}n}}, \tag{10}$$

其中 $n_g=N_g-1$ 且 $n=\sum_{g=1}^{q}n_g=N-q$. 分子和样本广义方差的几何加权平均的幂成正比, 分母和样本协方差阵的算术加权平均的行列式的幂成正比.

在两个样本的标量情形 $(p=1)$ 时, 准则 (10) 是

$$\frac{(n_1)^{\frac{1}{2}n_1}(n_2)^{\frac{1}{2}n_2}(s_1^2)^{\frac{1}{2}n_1}(s_2^2)^{\frac{1}{2}n_2}}{(n_1s_1^2+n_2s_2^2)^{\frac{1}{2}(n_1+n_2)}}=\frac{(n_1)^{\frac{1}{2}n_1}(n_2)^{\frac{1}{2}n_2}\cdot F^{\frac{1}{2}n_1}}{(n_1F+n_2)^{\frac{1}{2}(n_1+n_2)}}, \tag{11}$$

其中 s_1^2 和 s_2^2 是 σ_1^2 和 σ_2^2(两个总体的方差) 的一般无偏估计并且

$$F = \frac{s_1^2}{s_2^2}. \tag{12}$$

因此临界区域

$$V_1 \leqslant V_1(\varepsilon) \tag{13}$$

是基于自由度为 n_1 和 n_2 的 F 统计量的, 并且不等式 (13) 暗示了一个为临界区域选择 $F_1(\varepsilon)$ 和 $F_2(\varepsilon)$ 的特定方法

$$F \leqslant F_1(\varepsilon), \quad F \geqslant F_2(\varepsilon). \tag{14}$$

Brown(1939) 和 Scheffé (1942) 已经证明 (14) 式将产生一个无偏检验.

Bartlett 对于用 V_1 来替代 λ_1 的做法给出了一个直观的论述. 他指出, 如果 N_1 很小, A_1 在 λ_1 上加的权就会太大, 并且其他效应可能会丢失. Perlman(1980) 已经证明基于 V_1 的检验是无偏的.

如果假设

$$\mathrm{E}(\boldsymbol{X}_\alpha^{(g)}) = \boldsymbol{\beta}_g z_\alpha^{(g)}, \tag{15}$$

其中 $z_\alpha^{(g)}$ 包括 k_g 个分量, 并且如果可以估计出矩阵 $\boldsymbol{\beta}_g$, 定义

$$\boldsymbol{A}_g = \sum_{\alpha=1}^{N_g} (\boldsymbol{x}_\alpha^{(g)} - \hat{\boldsymbol{\beta}}_g z_\alpha^{(g)})(\boldsymbol{x}_\alpha^{(g)} - \hat{\boldsymbol{\beta}}_g z_\alpha^{(g)})', \tag{16}$$

则可以利用 (10) 式, 其中 $n_g = N_g - k_g$.

统计问题 (参数空间 Ω 和原假设 ω) 关于总体内位置的变化和普通的线性变换

$$\boldsymbol{X}^{*(g)} = \boldsymbol{C}\boldsymbol{X}^{(g)} + \boldsymbol{\nu}^{(g)}, \quad g = 1, \cdots, q, \tag{17}$$

是不变的, 其中 $\boldsymbol{C}$ 是非奇异的. 每个矩阵 $\boldsymbol{A}_g$ 在位置变化下是不变的, 并且调整后的准则 (10) 是不变的:

$$V_1^* = \frac{\prod_{g=1}^q |\boldsymbol{A}_g^*|^{\frac{1}{2}n_g}}{|\boldsymbol{A}^*|^{\frac{1}{2}n}} = \frac{\prod_{g=1}^q |\boldsymbol{C}\boldsymbol{A}_g\boldsymbol{C}'|^{\frac{1}{2}n_g}}{|\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}'|^{\frac{1}{2}n}} = \frac{\prod_{g=1}^q |\boldsymbol{A}_g|^{\frac{1}{2}n_g}}{|\boldsymbol{A}|^{\frac{1}{2}n}} = V_1. \tag{18}$$

同样, 似然比准则 (8) 是不变的.

另外一个不变检验过程 [Nagao(1973a)] 基于准则

$$\frac{1}{2}\sum_{g=1}^q n_g \operatorname{tr}(\boldsymbol{S}_g\boldsymbol{S}^{-1} - \boldsymbol{I})^2 = \frac{1}{2}\sum_{g=1}^q n_g \operatorname{tr}(\boldsymbol{S}_g - \boldsymbol{S})\boldsymbol{S}^{-1}(\boldsymbol{S}_g - \boldsymbol{S})\boldsymbol{S}^{-1}, \tag{19}$$

其中 $\boldsymbol{S}_g = (1/n_g)\boldsymbol{A}_g$ 并且 $\boldsymbol{S} = (1/n)\boldsymbol{A}$. (参见 7.8 节.)

10.3 检验几个正态分布相等的准则

在 8.8 节中我们考虑过均值向量相等的检验, 当时我们假设协方差阵是相等的, 即我们检验的是

$$H_2: \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(q)}, \quad 给定\ \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_q. \tag{1}$$

H_2 中假设的检验在 10.2 节中考虑过了. 现在让我们考虑均值和协方差都相等的假设, 即 H_1 和 H_2 相结合. 我们检验

$$H: \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)} = \cdots = \boldsymbol{\mu}^{(q)}, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_q. \tag{2}$$

正如 10.2 节中一样, 令 $\boldsymbol{x}_\alpha^{(g)}\,(\alpha = 1, \cdots, N_g)$ 是从 $N(\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g)\,(g = 1, \cdots, q)$ 中得到的观测. 则 Ω 是 $\{\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g\}\,(g = 1, \cdots, q)$ 的无限制的参数空间, 其中 $\boldsymbol{\Sigma}_g$ 是正定的, 而 ω^* 包含由 (2) 限制的空间.

似然函数由 10.2 节的 (3) 式给出. 10.2 节中的假设 H_1 是参数点在 ω 中; 8.8 节的假设 H_2 是, 如果参数点在 ω 中, 而 $\omega \supset \omega^*$, 则参数点在 ω^* 中; 这里的假设 H 是已知参数点在 Ω 中则它在 ω^* 中.

我们利用下面的引理.

引理 10.3.1 令 $\boldsymbol{y}$ 是密度为 $f(\boldsymbol{z}, \boldsymbol{\theta})$ 的随机向量的一个观测向量, 其中 $\boldsymbol{\theta}$ 是空间 Ω 中的一个参数向量. 令 H_a 为假设 "$\boldsymbol{\theta} \in \Omega_a \subset \Omega$", H_b 为假设 "给定 $\boldsymbol{\theta} \in \Omega_a$, $\boldsymbol{\theta} \in \Omega_b \subset \Omega_a$", 且令 H_{ab} 为假设 "给定 $\boldsymbol{\theta} \in \Omega$", $\boldsymbol{\theta} \in \Omega_b$. 如果检验 H_a 的似然比准则 λ_a、检验 H_b 的似然比准则 λ_b 和检验 H_{ab} 的似然比准则 λ_{ab} 分别是由观测向量 $\boldsymbol{y}$ 唯一定义的, 则

$$\lambda_{ab} = \lambda_a \lambda_b. \tag{3}$$

证明 引理可由下面的定义得出:

$$\lambda_a = \frac{\max_{\boldsymbol{\theta} \in \Omega_a} f(\boldsymbol{y}, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Omega} f(\boldsymbol{y}, \boldsymbol{\theta})}, \tag{4}$$

$$\lambda_b = \frac{\max_{\boldsymbol{\theta} \in \Omega_b} f(\boldsymbol{y}, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Omega_a} f(\boldsymbol{y}, \boldsymbol{\theta})}, \tag{5}$$

$$\lambda_{ab} = \frac{\max_{\boldsymbol{\theta} \in \Omega_b} f(\boldsymbol{y}, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Omega} f(\boldsymbol{y}, \boldsymbol{\theta})}. \quad \blacksquare \tag{6}$$

所以假设 H 的似然比准则是 H_1 和 H_2 的似然比准则的乘积,

$$\lambda = \lambda_1 \lambda_2 = \left(\prod_{g=1}^{q} \frac{|\boldsymbol{A}_g|^{\frac{1}{2}N_g}}{N_g^{\frac{1}{2}pN_g}} \right) \frac{N^{\frac{1}{2}pN}}{|\boldsymbol{B}|^{\frac{1}{2}N}}, \tag{7}$$

其中

$$\begin{aligned} \boldsymbol{B} &= \sum_{g=1}^{q} \sum_{\alpha=1}^{N_g} (\boldsymbol{x}_\alpha^{(g)} - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha^{(g)} - \bar{\boldsymbol{x}})' \\ &= \boldsymbol{A} + \sum_{g=1}^{q} N_g (\bar{\boldsymbol{x}}^{(g)} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)} - \bar{\boldsymbol{x}})'. \end{aligned} \tag{8}$$

临界区域定义如下:

$$\lambda \leqslant \lambda(\varepsilon), \tag{9}$$

其中 $\lambda(\varepsilon)$ 的选择是为了使在 H 下 (9) 式的概率为 ε.

令

$$V_2 = \frac{|\boldsymbol{A}|^{\frac{1}{2}n}}{|\boldsymbol{B}|^{\frac{1}{2}n}} = \lambda_2^{n/N}, \tag{10}$$

这和检验 H_2 的 λ_2 等价, 也就是 8.8 节中 (12) 式的 λ. 我们可能考虑

$$V = V_1 V_2 = \frac{\prod_{g=1}^{q} |\boldsymbol{A}_g|^{\frac{1}{2}n_g}}{|\boldsymbol{B}|^{\frac{1}{2}n}}. \tag{11}$$

但是, Perlman(1980) 已经证明过似然比检验是无偏的.

10.4 准则的分布

10.4.1 分布的特征

首先我们考虑 10.2 节中给出的 V_1. 如果

$$V_{1g} = \frac{|\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{g-1}|^{\frac{1}{2}(n_1+\cdots+n_{g-1})}|\boldsymbol{A}_g|^{\frac{1}{2}n_g}}{|\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_g|^{\frac{1}{2}(n_1+\cdots+n_g)}}, \quad g = 2, \cdots, q \tag{1}$$

则

$$V_1 = \prod_{g=2}^{q} V_{1g}. \tag{2}$$

定理 10.4.1 当 $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q$ 且 $n_g \geqslant p\,(g = 1, \cdots, q)$ 时, 由 (1) 式定义的 $V_{12}, V_{13}, \cdots, V_{1q}$ 是独立的.

这个定理可由下面引理得到.

引理 10.4.1 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是独立的, 且分别服从 $W(\boldsymbol{\Sigma}, m)$ 和 $W(\boldsymbol{\Sigma}, n)$, $n \geqslant p$, $m \geqslant p$, $\boldsymbol{C}$ 满足 $\boldsymbol{C}(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}' = \boldsymbol{I}$, 则 $\boldsymbol{A}+\boldsymbol{B}$ 和 $\boldsymbol{CAC}'$ 是独立地分布的; $\boldsymbol{A}+\boldsymbol{B}$ 服从自由度为 $m+n$ 的 Wishart 分布, $\boldsymbol{CAC}'$ 服从自由度为 n 和 m 的多元贝塔分布.

引理的证明 $\boldsymbol{D} = \boldsymbol{A}+\boldsymbol{B}$ 和 $\boldsymbol{E} = \boldsymbol{CAC}'$ 的密度是把 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的联合密度中的 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别替换为 $\boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1}$ 和 $\boldsymbol{D} - \boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1} = \boldsymbol{C}^{-1}(\boldsymbol{I}-\boldsymbol{E})\boldsymbol{C}'^{-1}$, 并且乘以雅可比行列式 $\mod |\boldsymbol{C}|^{-(p+1)} = |\boldsymbol{D}|^{\frac{1}{2}(p+1)}$, 对于正定的 $\boldsymbol{D}$, $\boldsymbol{E}$ 和 $\boldsymbol{I}-\boldsymbol{E}$, 得到

$$\begin{aligned}
&K(\boldsymbol{\Sigma}, m)K(\boldsymbol{\Sigma}, n)|\boldsymbol{C}^{-1}\boldsymbol{E}\boldsymbol{C}'^{-1}|^{\frac{1}{2}(m-p-1)} \\
&\cdot|\boldsymbol{C}^{-1}(\boldsymbol{I}-\boldsymbol{E})\boldsymbol{C}'^{-1}|^{\frac{1}{2}(n-p-1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{D}}|\boldsymbol{D}|^{\frac{1}{2}(p+1)} \\
=&K(\boldsymbol{\Sigma}, m+n)|\boldsymbol{D}|^{\frac{1}{2}(m+n-p-1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{D}} \\
&\cdot\frac{\Gamma_p\left[\frac{1}{2}(m+n)\right]}{\Gamma_p\left(\frac{1}{2}m\right)\Gamma_p\left(\frac{1}{2}n\right)}|\boldsymbol{E}|^{\frac{1}{2}(m-p-1)}|\boldsymbol{I}-\boldsymbol{E}|^{\frac{1}{2}(n-p-1)}.
\end{aligned} \tag{3}$$

■

定理的证明　如果我们令 $\boldsymbol{A}_1+\cdots+\boldsymbol{A}_g=\boldsymbol{D}_g$, $\boldsymbol{C}_g(\boldsymbol{A}_1+\cdots+\boldsymbol{A}_{g-1})\boldsymbol{C}_g'=\boldsymbol{E}_g$, 其中 $\boldsymbol{C}_g\boldsymbol{D}_g\boldsymbol{C}_g'=\boldsymbol{I}$, $g=2,\cdots,q$, 则

$$V_{1g}=\frac{|\boldsymbol{C}_g^{-1}\boldsymbol{E}_g\boldsymbol{C}_g'^{-1}|^{\frac{1}{2}(n_1+\cdots+n_{g-1})}|\boldsymbol{C}_g^{-1}(\boldsymbol{I}-\boldsymbol{E}_g)\boldsymbol{C}_g'^{-1}|^{\frac{1}{2}n_g}}{|\boldsymbol{C}_g^{-1}\boldsymbol{C}_g'^{-1}|^{\frac{1}{2}(n_1+\cdots+n_g)}} \tag{4}$$

$$=|\boldsymbol{E}_g|^{\frac{1}{2}(n_1+\cdots+n_{g-1})}|\boldsymbol{I}-\boldsymbol{E}_g|^{\frac{1}{2}n_g},\quad g=2,\cdots,q,$$

并且由引理 10.4.1 得 $\boldsymbol{E}_2,\cdots,\boldsymbol{E}_q$ 是独立的. ■

现在我们要找 V_{1g} 的分布的特征. 统计量 V_{1g} 有下面的形式

$$\frac{|\boldsymbol{B}|^b|\boldsymbol{C}|^c}{|\boldsymbol{B}+\boldsymbol{C}|^{b+c}}. \tag{5}$$

令 $\boldsymbol{B}_i$ 和 $\boldsymbol{C}_i$ 分别是矩阵 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 的阶数为 i 的左上角子方阵. 定义 $\boldsymbol{b}_{(i)}$ 和 $\boldsymbol{c}_{(i)}$ 为

$$\boldsymbol{B}_i=\begin{pmatrix}\boldsymbol{B}_{i-1} & \boldsymbol{b}_{(i)}\\ \boldsymbol{b}_{(i)}' & b_{ii}\end{pmatrix},\quad \boldsymbol{C}_i=\begin{pmatrix}\boldsymbol{C}_{i-1} & \boldsymbol{c}_{(i)}\\ \boldsymbol{c}_{(i)}' & c_{ii}\end{pmatrix},\quad i=2,\cdots,p. \tag{6}$$

则 (5) 式是 ($\boldsymbol{B}_0=\boldsymbol{C}_0=\boldsymbol{I}$, $\boldsymbol{b}_{(1)}=\boldsymbol{c}_{(1)}=\boldsymbol{0}$)

$$\frac{|\boldsymbol{B}|^b|\boldsymbol{C}|^c}{|\boldsymbol{B}+\boldsymbol{C}|^{b+c}}=\prod_{i=1}^{p}\frac{|\boldsymbol{B}_i|^b|\boldsymbol{C}_i|^c}{|\boldsymbol{B}_{i-1}|^b|\boldsymbol{C}_{i-1}|^c}\cdot\frac{|\boldsymbol{B}_{i-1}+\boldsymbol{C}_{i-1}|^{b+c}}{|\boldsymbol{B}_i+\boldsymbol{C}_i|^{b+c}} \tag{7}$$

$$=\prod_{i=1}^{p}\frac{(b_{ii}-\boldsymbol{b}_{(i)}'\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)})^b(c_{ii}-\boldsymbol{c}_{(i)}'\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)})^c}{\left[b_{ii}+c_{ii}-(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)})'(\boldsymbol{B}_{i-1}+\boldsymbol{C}_{i-1})^{-1}(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)})\right]^{b+c}}$$

$$=\prod_{i=1}^{p}\left\{\frac{b_{ii\cdot i-1}^b c_{ii\cdot i-1}^c}{(b_{ii\cdot i-1}+c_{ii\cdot i-1})^{b+c}}\cdot\frac{(b_{ii\cdot i-1}+c_{ii\cdot i-1})^{b+c}}{\left[\begin{array}{l}b_{ii\cdot i-1}+c_{ii\cdot i-1}+\boldsymbol{b}_{(i)}'\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)}'\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)}\\ -(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)})'(\boldsymbol{B}_{i-1}+\boldsymbol{C}_{i-1})^{-1}(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)})\end{array}\right]^{b+c}}\right\},$$

其中 $b_{ii\cdot i-1}=b_{ii}-\boldsymbol{b}_{(i)}'\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}$, $c_{ii\cdot i-1}=c_{ii}-\boldsymbol{c}_{(i)}'\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)}$. 对于 $i=1$, 第二项定义为 1.

现在要讨论的是当 $\boldsymbol{B}$ 和 $\boldsymbol{C}$ 是独立的且分别服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)$ 时, (7) 式右边的比是统计独立的. 由定理 4.3.3 可知, 对于固定的 $\boldsymbol{B}_{i-1}$, $\boldsymbol{b}_{(i)}$ 和 $b_{ii\cdot i-1}$ 是独立的, 且分别服从 $N(\boldsymbol{\beta}_{(i)},\sigma_{ii\cdot i-1}\boldsymbol{B}_{i-1}^{-1})$ 和自由度为 $m-(i-1)$ 的 $\sigma_{ii\cdot i-1}\chi^2$ 分布. 引理 10.4.1 表明第一项 (它是 $b_{ii\cdot i-1}/c_{ii\cdot i-1}$ 的函数) 和 $b_{ii\cdot i-1}+c_{ii\cdot i-1}$ 是独立的.

我们应用如下引理.

引理 10.4.2 对于正定的 $\boldsymbol{B}_{i-1}$ 和 $\boldsymbol{C}_{i-1}$,

$$\begin{aligned}&\boldsymbol{b}'_{(i)}\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}+\boldsymbol{c}'_{(i)}\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)}-(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)})'(\boldsymbol{B}_{i-1}+\boldsymbol{C}_{i-1})^{-1}(\boldsymbol{b}_{(i)}+\boldsymbol{c}_{(i)}) \qquad (8)\\&=(\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}-\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)})'(\boldsymbol{B}_{i-1}^{-1}+\boldsymbol{C}_{i-1}^{-1})^{-1}(\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}-\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)}).\end{aligned}$$

证明 利用 $(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}=[\boldsymbol{C}^{-1}(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{B}^{-1}]^{-1}=\boldsymbol{B}(\boldsymbol{B}+\boldsymbol{C})^{-1}\boldsymbol{C}$ 可证明 (8) 式的左边是 (省略 i 和 $i-1$)

$$\begin{aligned}&\boldsymbol{b}'\boldsymbol{B}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})\boldsymbol{b}+\boldsymbol{c}'(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{C}^{-1}\boldsymbol{c} \qquad (9)\\&-(\boldsymbol{b}+\boldsymbol{c})'\boldsymbol{B}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{C}^{-1}(\boldsymbol{b}+\boldsymbol{c})\\=&\boldsymbol{b}'\boldsymbol{B}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{B}^{-1}\boldsymbol{b}+\boldsymbol{c}'\boldsymbol{C}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{C}^{-1}\boldsymbol{c}\\&-\boldsymbol{b}'\boldsymbol{B}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})^{-1}\boldsymbol{C}^{-1}\boldsymbol{c}-\boldsymbol{c}'\boldsymbol{C}^{-1}(\boldsymbol{B}^{-1}+\boldsymbol{C}^{-1})\boldsymbol{B}^{-1}\boldsymbol{b}.\end{aligned}$$

这正是 (8) 式的右边. ■

(7) 式中第 i 个第二项的分母是其分子乘以 (8) 式. $\boldsymbol{B}_{i-1}^{-1}\boldsymbol{b}_{(i)}-\boldsymbol{C}_{i-1}^{-1}\boldsymbol{c}_{(i)}$ 的条件分布是均值为 $\boldsymbol{B}_{i-1}^{-1}\boldsymbol{\beta}_{(i)}-\boldsymbol{C}_{i-1}^{-1}\boldsymbol{\gamma}_{(i)}$ 协方差阵为 $\sigma_{ii\cdot i-1}(\boldsymbol{B}_{i-1}^{-1}+\boldsymbol{C}_{i-1}^{-1})$ 的正态分布. 协方差阵是 $\sigma_{ii\cdot i-1}$ 乘以 (8) 式右边第二个矩阵的逆. 因此 (8) 式服从自由度为 $i-1$ 的 $\sigma_{ii\cdot i-1}\chi^2$ 分布, 且与 $\boldsymbol{B}_{i-1}$, $\boldsymbol{C}_{i-1}$, $b_{ii\cdot i-1}$ 和 $c_{ii\cdot i-1}$ 独立.

则

$$\frac{b_{ii\cdot i-1}^{b}c_{ii\cdot i-1}^{c}}{(b_{ii\cdot i-1}+c_{ii\cdot i-1})^{b+c}}=\left(\frac{b_{ii\cdot i-1}}{b_{ii\cdot i-1}+c_{ii\cdot i-1}}\right)^{b}\left(\frac{c_{ii\cdot i-1}}{b_{ii\cdot i-1}+c_{ii\cdot i-1}}\right)^{c} \qquad (10)$$

和 $X_i^b(1-X_i)^c$ 分布相同, 其中 X_i 有分布 $\beta[\frac{1}{2}(m-i+1),\frac{1}{2}(n-i+1)]$, $i=1,\cdots,p$. 同样

$$\left[\frac{b_{ii\cdot i-1}+c_{ii\cdot i-1}}{b_{ii\cdot i-1}+c_{ii\cdot i-1}+(8)}\right]^{b+c},\quad i=2,\cdots,p, \qquad (11)$$

和 Y_i^{b+c} 分布相同, 其中 Y_i 有分布 $\beta\left[\frac{1}{2}(m+n)-i+1,\frac{1}{2}(i-1)\right]$. 故 (5) 和 $\prod_{i=1}^{p}X_i^b(1-X_i)^c\prod_{i=2}^{p}Y_i^{b+c}$ 同分布, 且这些因子之间是相互独立的.

定理 10.4.2

$$V_1=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}X_{ig}^{\frac{1}{2}(n_1+\cdots+n_{g-1})}(1-X_{ig})^{\frac{1}{2}n_g}\cdot\prod_{i=2}^{p}Y_{ig}^{\frac{1}{2}(n_1+\cdots+n_g)}\right\}, \qquad (12)$$

其中所有的 X 和 Y 是独立的, X_{ig} 有分布 $\beta\left[\frac{1}{2}(n_1+\cdots+n_{g-1}-i+1),\frac{1}{2}(n_g-i+1)\right]$, Y_{ig} 有分布 $\beta[\frac{1}{2}(n_1+\cdots+n_g)-i+1,\frac{1}{2}(i-1)]$.

证明 由定理 10.4.1 知因子 $V_{12},\cdots,V_{1q}$ 是独立的. 每一项 V_{1g} 按照 (7) 式分解, 并且因子是独立的. ■

V_1 的因子可以解释为子假设的检验准则. 依赖于 X_{i2} 的项是检验假设 "$\sigma_{ii\cdot i-1}^{(1)}=\sigma_{ii\cdot i-1}^{(2)}$" 的准则, 依赖于 Y_{i2} 的项是检验 "给定 $\sigma_{ii\cdot i-1}^{(1)}=\sigma_{ii\cdot i-1}^{(2)}$ 和 $\boldsymbol{\Sigma}_{i-1,1}=\boldsymbol{\Sigma}_{i-1,2}$ 时 $\sigma_{(i)}^{(1)}=\sigma_{(i)}^{(2)}$" 的准则. 类似地, 依赖于 X_{ig} 和 Y_{ig} 的项提供了检验 "给定 $\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_{g-1}$ 时 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_g$" 的准则.

现在考虑 10.3 节 (7) 式给出的检验假设 "$\boldsymbol{\mu}^{(1)}=\cdots=\boldsymbol{\mu}^{(q)}$ 和 $\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_q$" 的似然比准则 λ. 它等价于准则

$$W=\frac{\prod_{g=1}^{q}|\boldsymbol{A}_g|^{\frac{1}{2}N_g}}{|\boldsymbol{A}_1+\cdots+\boldsymbol{A}_q|^{\frac{1}{2}(N_1+\cdots+N_g)}}\cdot\frac{|\boldsymbol{A}_1+\cdots+\boldsymbol{A}_q|^{\frac{1}{2}N}}{|\boldsymbol{A}_1+\cdots+\boldsymbol{A}_q+\sum_{g=1}^{q}N_g(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'|^{\frac{1}{2}N}}.\tag{13}$$

(13) 式的两个因子是独立的, 因为第一个因子与 $\boldsymbol{A}_1+\cdots+\boldsymbol{A}_q$(由引理 10.4.1 和定理 10.4.1 的证明) 和 $\bar{\boldsymbol{x}}^{(1)},\cdots,\bar{\boldsymbol{x}}^{(q)}$ 是独立的.

定理 10.4.3

$$W=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}X_{ig}^{\frac{1}{2}(N_1+\cdots+N_{g-1})}(1-X_{ig})^{\frac{1}{2}N_g}\prod_{i=2}^{p}Y_{ig}^{\frac{1}{2}(N_1+\cdots+N_g)}\right\}\prod_{i=1}^{p}Z_i^{\frac{1}{2}N},\tag{14}$$

其中所有的 X, Y 和 Z 是独立的, X_{ig} 有分布 $\beta\left[\frac{1}{2}(n_1+\cdots+n_{g-1}-i+1),\frac{1}{2}(n_g-i+1)\right]$, Y_{ig} 有分布 $\beta\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1,\frac{1}{2}(i-1)\right]$, Z_i 有分布 $\beta\left[\frac{1}{2}(n+1-i),\frac{1}{2}(q-1)\right]$.

证明　(13) 式的第一个因子的特征对应于 V_1 的特征, 只是把 X_{ig} 和 $1-X_{ig}$ 的指数由 n_g 改为 N_g. 第二项是 $U_{p,q-1,n}^{\frac{1}{2}N}$, 其特征可由定理 8.4.1 得到. ■

10.4.2　分布的矩

我们现在要求 V_1 和 W 的矩. 由于 $0\leqslant V_1\leqslant 1$ 和 $0\leqslant W\leqslant 1$, 所以矩可以唯一确定分布. V_1 的 h 阶矩可以由定理 10.4.2 中的分布的特征求出:

$$\begin{aligned}\mathrm{E}(V_1^h)&=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\mathrm{E}\left(X_{ig}^{\frac{1}{2}(n_1+\cdots+n_{g-1})h}(1-X_{ig})^{\frac{1}{2}n_gh}\right)\prod_{i=2}^{p}\mathrm{E}\left(Y_{ig}^{\frac{1}{2}(n_1+\cdots+n_g)h}\right)\right\}\\&=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1})(1+h)-\frac{1}{2}(i-1)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}-i+1)\right]}\right.\\&\quad\cdot\frac{\Gamma\left[\frac{1}{2}n_g(1+h)-\frac{1}{2}(i-1)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]}{\Gamma\left[\frac{1}{2}(n_g-i+1)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-i+1\right]}\\&\quad\left.\cdot\prod_{i=2}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-i+1\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g-i+1)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)(1+h)-\frac{1}{2}(i-1)\right]}\right\}\end{aligned}\tag{15}$$

$$=\prod_{i=1}^{p}\left\{\frac{\Gamma\left[\frac{1}{2}(n+1-i)\right]}{\Gamma\left[\frac{1}{2}(n+hn+1-i)\right]}\prod_{g=1}^{q}\frac{\Gamma\left[\frac{1}{2}(n_g+hn_g+1-i)\right]}{\Gamma\left[\frac{1}{2}(n_g+1-i)\right]}\right\}$$

$$=\frac{\Gamma_p\left(\frac{1}{2}n\right)}{\Gamma_p\left[\frac{1}{2}(n+hn)\right]}\prod_{g=1}^{q}\frac{\Gamma_p\left[\frac{1}{2}(n_g+hn_g)\right]}{\Gamma_p\left(\frac{1}{2}n_g\right)}.$$

W 的 h 阶矩可以由定理 10.4.3 中它的表示求得. 我们有

$$\mathrm{E}(W^h)=\prod_{g=2}^{q}\prod_{i=1}^{p}\mathrm{E}\left(X_{ig}^{\frac{1}{2}(N_1+\cdots+N_{g-1})h}(1-X_{ig})^{\frac{1}{2}N_gh}\right) \tag{16}$$

$$\cdot\prod_{i=2}^{p}\mathrm{E}\left(Y_{ig}^{\frac{1}{2}(N_1+\cdots+N_g)h}\right)\mathrm{E}\left(U_{p,q-1,n}^{\frac{1}{2}Nh}\right)$$

$$=\prod_{g=2}^{q}\left\{\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}+1-i)+\frac{1}{2}h(N_1+\cdots+N_{g-1})\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_{g-1}+1-i)\right]\Gamma\left[\frac{1}{2}(n_g+1-i)\right]}\right.$$

$$\cdot\frac{\Gamma\left[\frac{1}{2}(n_g+1-i+N_gh)\right]\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)-i+1\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+\frac{1}{2}h(N_1+\cdots+N_g)+1-i\right]}$$

$$\cdot\prod_{i=2}^{p}\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+\frac{1}{2}h(N_1+\cdots+N_g)+1-i\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g)+1-i\right]}$$

$$\left.\cdot\frac{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g+1-i)\right]}{\Gamma\left[\frac{1}{2}(n_1+\cdots+n_g+1-i)+\frac{1}{2}h(N_1+\cdots+N_g)\right]}\right\}$$

$$\cdot\prod_{i=1}^{p}\frac{\Gamma\left[\frac{1}{2}(n+1-i+hN)\right]\Gamma\left[\frac{1}{2}(N-i)\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]\Gamma\left[\frac{1}{2}(N+hN-i)\right]}$$

$$=\prod_{i=1}^{p}\left\{\prod_{g=1}^{q}\frac{\Gamma\left[\frac{1}{2}(N_g+hN_g-i)\right]}{\Gamma\left[\frac{1}{2}(N_g-i)\right]}\right\}\frac{\Gamma\left[\frac{1}{2}(N-i)\right]}{\Gamma\left[\frac{1}{2}(N+hN-i)\right]}$$

$$=\frac{\Gamma_p\left(\frac{1}{2}n\right)}{\Gamma_p\left(\frac{1}{2}n+\frac{1}{2}hN\right)}\prod_{g=1}^{q}\frac{\Gamma_p\left[\frac{1}{2}(n_g+hN_g)\right]}{\Gamma_p\left(\frac{1}{2}n_g\right)}.$$

我们总结得到如下定理.

定理 10.4.4 令 V_1 为 10.2 节中 (10) 式定义的检验假设 "$H_1:\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_q$" 的准则, 其中 $\boldsymbol{A}_g$ 是 n_g 乘以样本协方差阵, 而 n_g+1 是从第 g 个总体中抽出的样本的大小; 令 W 为 (13) 式定义的检验假设 "$H:\boldsymbol{\mu}_1=\cdots=\boldsymbol{\mu}_q$" 和 H_1 的准则, 其中 $\boldsymbol{B}=\boldsymbol{A}+\sum_g N_g(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'$. 当 H_1 为真时 V_1 的 h 阶矩由 (15) 式给出. W 的 h 阶矩是检验 H 的准则, 它由 (16) 式给出.

这个定理首先由 Wilks(1932) 证明. 另一种方法可以参见习题 10.5.

如果 p 是偶数, 即 $p=2r$, 我们可以应用伽玛函数的倍量公式 $[\Gamma(\alpha+\frac{1}{2})\Gamma(\alpha+$

$1)=\sqrt{\pi}\Gamma(2\alpha+1)2^{-2\alpha}]$. 则

$$\mathrm{E}(V_1^h)=\prod_{j=1}^{r}\left\{\left[\prod_{g=1}^{q}\frac{\Gamma(n_g+hn_g+1-2j)}{\Gamma(n_g+1-2j)}\right]\frac{\Gamma(n+1-2j)}{\Gamma(n+hn+1-2j)}\right\}, \tag{17}$$

$$\mathrm{E}(W^h)=\prod_{j=1}^{r}\left\{\left[\prod_{g=1}^{q}\frac{\Gamma(n_g+hN_g+1-2j)}{\Gamma(n_g+1-2j)}\right]\frac{\Gamma(N-2j)}{\Gamma(N+hN-2j)}\right\}. \tag{18}$$

原则上说, 可以通过积分这些因子的分布来得到 V_1 和 W 的分布. 在 10.6 节中我们考虑 $p=2, q=2$ 时 V_1 的分布 ($p=1, q=2$ 时是 F 统计量的函数). 在其他情况下, 积分求不出来. 下节中我们将利用渐近展开来得到概率. Box(1949) 给出了其他一些近似分布.

10.4.3 逐步下降检验

关于独立因子的准则的分布特征表明, 检验假设 H_1 和 H 可以由依次检验分量假设来完成. 首先, 我们考虑 $q=2$ 时的检验 $H_1: \boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$. 对于 $i=2,\cdots,p, g=1,2$, 令

$$\boldsymbol{X}_{(i)}^{(g)}=\begin{pmatrix}\boldsymbol{X}_{(i-1)}^{(g)}\\ X_i^{(g)}\end{pmatrix},\quad \boldsymbol{\mu}_{(i)}^{(g)}=\begin{pmatrix}\boldsymbol{\mu}_{(i-1)}^{(g)}\\ \mu_i^{(g)}\end{pmatrix},\quad \boldsymbol{\Sigma}_i^{(g)}=\begin{bmatrix}\boldsymbol{\Sigma}_{(i-1)}^{(g)} & \boldsymbol{\sigma}_{(i)}^{(g)}\\ \boldsymbol{\sigma}_{(i)}^{(g)\prime} & \sigma_{ii}^{(g)}\end{bmatrix}, \tag{19}$$

给定 $\boldsymbol{X}_{(i-1)}^{(g)}=\boldsymbol{x}_{(i-1)}^{(g)}$ 时 $X_i^{(g)}$ 的条件分布为

$$N\left[\mu_i^{(g)}+\boldsymbol{\sigma}_{(i)}^{(g)\prime}\boldsymbol{\Sigma}_{i-1}^{-1}(\boldsymbol{x}_{(i-1)}^{(g)}-\boldsymbol{\mu}_{(i-1)}^{(g)}),\sigma_{ii\cdot i-1}^{(g)}\right], \tag{20}$$

其中 $\sigma_{ii\cdot i-1}^{(g)}=\sigma_{ii}^{(g)}-\boldsymbol{\sigma}_{(i)}^{(g)\prime}\boldsymbol{\Sigma}_{i-1}^{(g)}\boldsymbol{\sigma}_{(i)}^{(g)}$. 假设 $\boldsymbol{X}$ 的分量按照重要性递减的顺序排列. 在第 i 步中分量假设 $\sigma_{ii\cdot i-1}^{(1)}=\sigma_{ii\cdot i-1}^{(2)}$ 由基于 $s_{ii\cdot i-1}^{(1)}/s_{ii\cdot i-1}^{(2)}$ 的 F 检验在显著性水平 ε_i 下检验. $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 像 $\boldsymbol{\Sigma}^{(1)}$ 和 $\boldsymbol{\Sigma}^{(2)}$ 一样分块. 如果这个假设被接受了, 则在 $\boldsymbol{\Sigma}_{i-1}^{(1)}=\boldsymbol{\Sigma}_{i-1}^{(2)}$(先前接受的假设) 下检验假设 $\boldsymbol{\sigma}_{(i)}^{(1)}=\boldsymbol{\sigma}_{(i)}^{(2)}$(或者 $\boldsymbol{\Sigma}_{i-1}^{(1)-1}\boldsymbol{\sigma}_{(i)}^{(1)}=\boldsymbol{\Sigma}_{i-1}^{(2)-1}\boldsymbol{\sigma}_{(1)}^{(2)}$) 的显著性水平为 δ_i. 准则为

$$\frac{\left(\boldsymbol{S}_{i-1}^{(1)-1}\boldsymbol{s}_{(i)}^{(1)}-\boldsymbol{S}_{i-1}^{(2)-1}\boldsymbol{s}_{(i)}^{(2)}\right)'\left(\boldsymbol{S}_{i-1}^{(1)-1}-\boldsymbol{S}_{i-1}^{(2)-1}\right)^{-1}\left(\boldsymbol{S}_{i-1}^{(1)-1}\boldsymbol{s}_{(i)}^{(1)}-\boldsymbol{S}_{i-1}^{(2)-1}\boldsymbol{s}_{(i)}^{(2)}\right)}{(i-1)s_{ii\cdot i-1}}, \tag{21}$$

其中 $(n_1+n_2-2i+2)s_{ii\cdot i-1}=(n_i-i+1)s_{ii\cdot i-1}^{(1)}+(n_2-i+1)s_{ii\cdot i-1}^{(2)}$. 在原假设下 (21) 有自由度为 $i-1$ 和 n_1+n_2-2i+2 的 F 分布. 如果接受了这个假设, 则做第 $i+1$ 步. 如果 $2p-1$ 个分量假设都被接受了, 则接受总的假设 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$. (在第一步中, $\boldsymbol{\sigma}_{(1)}^{(g)}$ 是空的.) 总的显著性水平是

$$1-\prod_{i=1}^{p}(1-\varepsilon_i)\prod_{i=2}^{p}(1-\delta_i). \tag{22}$$

如果有一个分量假设被拒绝了, 则拒绝总假设.

如果 $q > 2$, 则原假设 $H_1 : \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q$ 被分为一列假设 $[1/(g-1)](\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_{g-1}) = \boldsymbol{\Sigma}_g$, 且依次检验. 每一个这样的矩阵假设都可以像检验 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ 一样检, 只是把 $\boldsymbol{S}_2$ 换为 $\boldsymbol{S}_g$, $\boldsymbol{S}_1$ 换为 $[1/(n_1 + \cdots + n_{g-1})](\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_{g-1})$.

对于假设 H 的情形, 首先考虑 $q = 2, \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, 且 $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$. 可以检验 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. 检验 $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(2)}$ 的步骤包括对于 $\mu_i^{(1)} = \mu_i^{(2)}$ 的 t 检验, 而后者是基于给定 $\boldsymbol{x}_{(i-1)}^{(1)}$ 和 $\boldsymbol{x}_{(i-1)}^{(2)}$ 下 $X_i^{(1)}$ 和 $X_i^{(2)}$ 的条件分布的. 另外也可以依次检验给定 $\boldsymbol{x}_{(i-1)}^{(1)}$ 和 $\boldsymbol{x}_{(i-1)}^{(2)}$ 下 $X_i^{(1)}$ 和 $X_i^{(2)}$ 的条件分布相等.

对于 $q > 2$, 可以检验假设 $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q$, 然后检验 $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q$. 相应地, 可以检验 $[1/(g-1)](\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_{g-1}) = \boldsymbol{\Sigma}_g$ 和 $[1/(g-1)](\boldsymbol{\mu}^{(1)} + \cdots + \boldsymbol{\mu}^{(g-1)}) = \boldsymbol{\mu}^{(g)}$.

10.5 准则的分布的渐近展开

我们再次利用定理 8.5.1 来得到 V_1 和 λ 的分布的渐近展开. 我们假设 $n_g = k_g n$, 其中 $\sum_{g=1}^q k_g = 1$. 渐近展开是通过当固定 $k_1, \cdots, k_g$ 时增加 n 来完成的. (我们可以只假设 $\lim n_g/n = k_g > 0$.)

$$\lambda_1^* = V_1 \cdot \frac{n^{\frac{1}{2}pn}}{\prod_{g=1}^q n_g^{\frac{1}{2}pn_g}} = V_1 \cdot \prod_{g=1}^q \left(\frac{n}{n_g}\right)^{\frac{1}{2}pn_g} = \left[\prod_{g=1}^q \left(\frac{1}{k_g}\right)^{k_g}\right]^{\frac{1}{2}pn} V_1 \tag{1}$$

的 h 阶矩是

$$\mathrm{E}(\lambda_1^{*h}) = K\left(\frac{\prod_{j=1}^p (\frac{1}{2}n)^{\frac{1}{2}n}}{\prod_{g=1}^q \prod_{i=1}^p (\frac{1}{2}n_g)^{\frac{1}{2}n_g}}\right)^h \frac{\prod_{g=1}^q \prod_{i=1}^p \Gamma\left[\frac{1}{2}n_g(1+h) + \frac{1}{2}(1-i)\right]}{\prod_{j=1}^p \Gamma\left[\frac{1}{2}n(1+h) + \frac{1}{2}(1-j)\right]}. \tag{2}$$

这和 8.6 节中 (1) 式的形式相同, 其中

$$\begin{aligned}
&b = p, \quad y_j = \frac{1}{2}n, \quad \eta_j = \frac{1}{2}(1-j), && j = 1, \cdots, p, \\
&a = pq, \quad x_k = \frac{1}{2}n_g, \quad k = (g-1)p+1, \cdots, gp, && g = 1, \cdots, q, \\
&\xi_k = \frac{1}{2}(1-i), \quad k = i, p+i, \cdots, (q-1)p+i, && i = 1, \cdots, p.
\end{aligned} \tag{3}$$

则

$$\begin{aligned}
f &= -2\left[\sum \xi_k - \sum \eta_i - \frac{1}{2}(a-b)\right] \\
&= -\left[q\sum_{i=1}^p (1-i) - \sum_{j=1}^p (1-j) - (qp-p)\right] \\
&= -\left[-\frac{1}{2}qp(p-1) + \frac{1}{2}p(p-1) - (q-1)p\right] = \frac{1}{2}(q-1)p(p+1),
\end{aligned} \tag{4}$$

$\varepsilon_j = \frac{1}{2}(1-p)n$, $j=1,\cdots,p$, 且 $\beta_k = \frac{1}{2}(1-p)n_g = \frac{1}{2}(1-p)k_g n$, $k=(g-1)p+1,\cdots,gp$.

为了消去第二项, 我们取 ρ 为

$$\rho = 1 - \left(\sum_{g=1}^{q}\frac{1}{n_g} - \frac{1}{n}\right)\frac{2p^2+3p-1}{6(p+1)(q-1)}. \tag{5}$$

则

$$\omega_2 = \frac{p(p+1)\left[(p-1)(p+2)\left(\sum_{g=1}^{q}\frac{1}{n_g^2}-\frac{1}{n^2}\right)-6(q-1)(1-\rho)^2\right]}{48\rho^2}. \tag{6}$$

因此

$$\begin{aligned}&\Pr\{-2\rho\ln\lambda_1^*\leqslant z\}\\ =&\Pr\{\chi_f^2\leqslant z\}+\omega_2\left[\Pr\{\chi_{f+4}^2\leqslant z\}-\Pr\{\chi_f^2\leqslant z\}\right]+O(n^{-3}).\end{aligned} \tag{7}$$

令 $\lambda = WN^{\frac{1}{2}pN}\prod_{g=1}^{q}N_g^{-\frac{1}{2}pN_g}$. 其 h 阶矩为

$$\mathrm{E}(\lambda^h) = K\left[\frac{\prod_{j=1}^{p}(\frac{1}{2}N)^{\frac{1}{2}N}}{\prod_{g=1}^{q}\prod_{i=1}^{p}(\frac{1}{2}N_g)^{\frac{1}{2}N_g}}\right]\frac{\prod_{g=1}^{q}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}N_g(1+h)-\frac{1}{2}i\right]}{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}N(1+h)-\frac{1}{2}j\right]}. \tag{8}$$

这和 8.5 节中 (1) 式的形式相同, 其中

$$\begin{aligned}&b=p, \quad y_j=\frac{1}{2}N=\frac{1}{2}\sum_{g=1}^{q}N_g, \quad \eta_j=-\frac{1}{2}j, \quad j=1,\cdots,p,\\ &a=pq, \quad x_k=\frac{1}{2}N_g, \quad k=(g-1)p+1,\cdots,gp, \quad g=1,\cdots,q,\\ &\xi_k=-\frac{1}{2}i, \quad k=i,p+i,\cdots,(q-1)p+i, \quad i=i,\cdots,p.\end{aligned} \tag{9}$$

自由度的基本数是 $f=\frac{1}{2}p(p+3)(q-1)$. 我们利用 8.5 节的 (11) 式, 其中 $\beta_k = (1-\rho)x_k$, $\varepsilon_j=(1-\rho)y_j$. 为了使 $\omega_1=0$, 我们取

$$\rho = 1-\left(\sum_{g=1}^{q}\frac{1}{N_g}-\frac{1}{N}\right)\frac{2p^2+9p+11}{6(q-1)(p+3)}. \tag{10}$$

则

$$\omega_2 = \frac{p(p+3)}{48\rho^2}\left[\sum_{g=1}^{q}\left(\frac{1}{N_g^2}-\frac{1}{N^2}\right)(p+1)(p+2)-6(1-\rho)^2(q-1)\right]. \tag{11}$$

$-2\rho\ln\lambda$ 的分布的渐近展开是

$$\begin{aligned}&\Pr\{-2\rho\ln\lambda\leqslant z\}\\ =&\Pr\{\chi_f^2\leqslant z\}+\omega_2\left[\Pr\{\chi_{f+4}^2\leqslant z\}-\Pr\{\chi_f^2\leqslant z\}\right]+O(n^{-3}).\end{aligned} \tag{12}$$

Box(1949) 详细考虑过 λ_1^* 的情形. 除了这个展开他还考虑了 8.6 节中 (13) 式的应用. 他还给出了 F 近似.

我们来看 E. S. Pearson and Wilks(1933) 给出的一个例子. 测量的是压铸铝的抗张强度 (X_1) 和硬度 (X_2). 五个样本中每一个都有 12 个观测. 五个样本的观测平方和与向量积是

$$\boldsymbol{A}_1=\begin{pmatrix}78.948 & 214.18\\214.18 & 1247.18\end{pmatrix}$$
$$\boldsymbol{A}_2=\begin{pmatrix}223.695 & 657.62\\657.62 & 2519.31\end{pmatrix}$$
$$\boldsymbol{A}_3=\begin{pmatrix}57.448 & 190.63\\190.63 & 1241.78\end{pmatrix}\tag{13}$$
$$\boldsymbol{A}_4=\begin{pmatrix}187.618 & 375.91\\375.91 & 1473.44\end{pmatrix}$$
$$\boldsymbol{A}_5=\begin{pmatrix}88.456 & 259.18\\259.18 & 1171.73\end{pmatrix}\tag{14}$$

它们的和为

$$\sum\boldsymbol{A}_i=\begin{pmatrix}636.165 & 1697.52\\1697.52 & 7653.44\end{pmatrix}.\tag{15}$$

$-\ln\lambda_1^*$ 是 5.399. 为了利用渐近展开我们得到 $\rho=152/165=0.9212$, $\omega_2=0.0022$. 由于 ω_2 很小, 我们可以认为 $-2\rho\ln\lambda_1^*$ 服从自由度为 12 的 χ^2 分布. 所以我们的观测准则显然不显著.

表 B.5[由 Korin(1969) 给出] 给出了 $-2\ln\lambda_1^*$ 的 5% 分位数, 其中 $p=2(1)6$, N_q 值很小, 且对于不同的 q 均有 $N_1=\cdots=N_q$.

10.1 节中准则 (19) 的极限分布也是 χ_f^2. 这个分布的渐近展开由 Nagao(1937b) 给出, 计算到 $1/n$ 阶, 包括自由度为 $f, f+2, f+4, f+6$ 的 χ^2 分布.

10.6 两个总体的情形

10.6.1 不变检验

当 $q=2$ 时, 原假设 H_1 是 $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$. 它对于下面的变换是不变的

$$\boldsymbol{x}^{*(1)}=\boldsymbol{C}\boldsymbol{x}^{(1)}+\boldsymbol{\nu}^{(1)},\quad \boldsymbol{x}^{*(2)}=\boldsymbol{C}\boldsymbol{x}^{(2)}+\boldsymbol{\nu}^{(2)},\tag{1}$$

其中 $\boldsymbol{C}$ 是非奇异的. 在位置变换 ($\boldsymbol{C}=\boldsymbol{I}$) 下参数的最大不变量是两个协方差阵 $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, 充分统计量 $\bar{\boldsymbol{x}}^{(1)}$, $\boldsymbol{S}_1$, $\bar{\boldsymbol{x}}^{(2)}$, $\boldsymbol{S}_2$ 的最大不变量是一对矩阵 $\boldsymbol{S}_1$, $\boldsymbol{S}_2$(或等价地 $\boldsymbol{A}_1$, $\boldsymbol{A}_2$). 变换 (1) 引出变换 $\boldsymbol{\Sigma}_1^*=\boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}'$, $\boldsymbol{\Sigma}_2^*=\boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}'$, $\boldsymbol{S}_1^*=\boldsymbol{C}\boldsymbol{S}_1\boldsymbol{C}'$, $\boldsymbol{S}_2^*=\boldsymbol{C}\boldsymbol{S}_2\boldsymbol{C}'$.

$$|\boldsymbol{\Sigma}_1-\lambda\boldsymbol{\Sigma}_2|=0\tag{2}$$

的根 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$ 在这些变换下是不变的, 因为

$$|\boldsymbol{\Sigma}_1^* - \lambda\boldsymbol{\Sigma}_2^*| = |\boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}' - \lambda\boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}'| = |\boldsymbol{C}\boldsymbol{C}'|\cdot|\boldsymbol{\Sigma}_1 - \lambda\boldsymbol{\Sigma}_2|. \tag{3}$$

而且, 这些根是唯一的不变量, 因为存在非奇异矩阵 $\boldsymbol{C}$ 使得

$$\boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}' = \boldsymbol{\Lambda}, \quad \boldsymbol{C}\boldsymbol{\Sigma}_2\boldsymbol{C}' = \boldsymbol{I}, \tag{4}$$

其中 $\boldsymbol{\Lambda}$ 是对角矩阵且其第 i 个对角元素为 λ_i, $i = 1,\cdots,p$. (参见附录的定理 A.2.2.) 类似地, $\boldsymbol{S}_1$ 和 $\boldsymbol{S}_2$ 的最大不变量是

$$|\boldsymbol{S}_1 - l\boldsymbol{S}_2| = 0 \tag{5}$$

的根 $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_p$.

定理 10.6.1 *在变换 (1) 下 $N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Sigma}_1)$ 和 $N(\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_2)$ 的参数的最大不变量是 (2) 的根 $\lambda_1 \geqslant \cdots \geqslant \lambda_p$. 充分统计量 $\bar{\boldsymbol{x}}^{(1)}$, $\boldsymbol{S}_1$,$\bar{\boldsymbol{x}}^{(2)}$, $\boldsymbol{S}_2$ 的最大不变量是 (5) 的根 $l_1 \geqslant \cdots \geqslant l_p$.*

任何不变检验准则可以由根 $l_1,\cdots,l_p$ 来表示. 准则 V_1 是 $n_1^{\frac{1}{2}pn_1}n_2^{\frac{1}{2}pn_2}$ 乘以

$$\frac{|\boldsymbol{S}_1|^{\frac{1}{2}n_1}|\boldsymbol{S}_2|^{\frac{1}{2}n_2}}{|n_1\boldsymbol{S}_1 + n_2\boldsymbol{S}_2|^{\frac{1}{2}n}} = \frac{|\boldsymbol{L}|^{\frac{1}{2}n_1}|\boldsymbol{I}|^{\frac{1}{2}n_2}}{|n_1\boldsymbol{L} + n_2\boldsymbol{I}|^{\frac{1}{2}n}} = \prod_{i=1}^{p}\frac{l_i^{\frac{1}{2}n_1}}{(n_1l_i + n_2)^{\frac{1}{2}n}}, \tag{6}$$

其中 $\boldsymbol{L}$ 是对角矩阵且第 i 个对角元素为 l_i. 如果较小的根太小或较大的根太大或两者均发生, 则拒绝原假设.

原假设是 $\lambda_1 = \cdots = \lambda_p = 1$. 原假设的任意有用的不变检验都有在空间 $l_1,\cdots,l_p$ 中的拒绝区域, 其包括在某种意义下远离 $l_1 = \cdots = l_p = 1$ 的点. 不变检验的功效通过根 $\lambda_1,\cdots,\lambda_p$ 依赖于参数.

10.2 节的准则 (19) 是 (其中 $n\boldsymbol{S} = n_1\boldsymbol{S}_1 + n_2\boldsymbol{S}_2$)

$$\begin{aligned}
&\frac{1}{2}n_1\,\mathrm{tr}\,\left[(\boldsymbol{S}_1 - \boldsymbol{S})\boldsymbol{S}^{-1}\right]^2 + \frac{1}{2}n_2\,\mathrm{tr}\,\left[(\boldsymbol{S}_2 - \boldsymbol{S})\boldsymbol{S}^{-1}\right]^2 \\
=&\frac{1}{2}n_1\,\mathrm{tr}\,\left[\boldsymbol{C}(\boldsymbol{S}_1 - \boldsymbol{S})\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}')^{-1}\right]^2 \\
&+\frac{1}{2}n_2\,\mathrm{tr}\,\left[\boldsymbol{C}(\boldsymbol{S}_2 - \boldsymbol{S})\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}')^{-1}\right]^2 \\
=&\frac{1}{2}n_1\,\mathrm{tr}\,\left[\left\{\boldsymbol{L} - \left(\frac{n_1}{n}\boldsymbol{L} + \frac{n_2}{n}\boldsymbol{I}\right)\right\}\left(\frac{n_1}{n}\boldsymbol{L} + \frac{n_2}{n}\boldsymbol{I}\right)^{-1}\right]^2 \\
&+\frac{1}{2}n_2\,\mathrm{tr}\,\left[\left\{\boldsymbol{L} - \left(\frac{n_1}{n}\boldsymbol{L} + \frac{n_2}{n}\boldsymbol{I}\right)\right\}\left(\frac{n_1}{n}\boldsymbol{L} + \frac{n_2}{n}\boldsymbol{I}\right)^{-1}\right]^2 \\
=&\frac{1}{2}n_1n_2n\sum_{i=1}^{p}\frac{(l_i - 1)^2}{(n_1l_i + n_2)^2}.
\end{aligned} \tag{7}$$

这个准则是度量 $l_1,\cdots,l_p$ 和 1 接近程度的, 如果它太大的话则拒绝假设. 在原假设下, 当 $n_1 \to \infty$, $n_2 \to \infty$ 且 n_1/n_2 趋于一个正的常数时, (7) 有自由度为

$f = \frac{1}{2}p(p+1)$ 的 χ^2 分布. Nagao(1973b) 给出了这个分布的到 $1/n$ 阶项的渐近展开.

Roy(1953) 给出了一个基于最大根 l_1 和最小根 l_p 的检验. 这个过程在 $l_1 > k_1$ 或 $l_p < k_p$ 时拒绝原假设, 其中 k_1 和 k_p 使得当 $\boldsymbol{\Lambda} = \boldsymbol{I}$ 时拒绝的概率达到期望的显著性水平. Roy(1957) 提出确定 k_1 和 k_p 使得检验是局部无偏的, 即功效函数在 $\boldsymbol{\Lambda} = \boldsymbol{I}$ 时有一个相对极小值. 由于在这个基础上很难确定 k_1 和 k_p, 所以需要给出其他方法. 极限 k_1 可以这样确定, 它使得 $\Pr(l_1 > k_1|H_1)$ 是显著性水平的一半, 或者 $\Pr(l_p < k_p|H_1)$ 是显著性水平的一半, 或者 $k_1 + k_p = 2$, 或者 $k_1k_p = 1$. 原则上, k_1 和 k_p 可以由 13.2 节给出的根的分布来确定. Schuurmann, Waikar, and Krishnaiah (1975) 和 Chu and Pillai(1979) 对于小的 p 值给出了 k_1 和 k_p 的精确值, Chu and Pillai(1970) 还对几个检验过程进行了功效对比.

在 $p=1$ 时, 充分统计量的唯一不变量是 S_1/S_2, 这正是自由度为 n_1 和 n_2 的 F 统计量. 准则 V_1 是 $(A_1/A_2)^{\frac{1}{2}n_1}[1 + A_1/A_2]^{-\frac{1}{2}n}$, V_1 小于一个常数的临界区域等于 F 统计量的双边临界区域. $n(B-A)/A$ 有独立的自由度为 1 和 n 的 F 分布. (参见 10.3 节.)

对于 $p=2$ 的情况, 由 10.4 节的 (15) 式可以得到 V_1 的 h 阶矩

$$\begin{aligned} \mathrm{E}(V_1^h) &= \frac{\Gamma(n_1 + hn_1 - 1)\Gamma(n_2 + hn_2 - 1)\Gamma(n-1)}{\Gamma(n_1 - 1)\Gamma(n_2 - 1)\Gamma(n + hn - 1)} \\ &= \mathrm{E}\left[X_1^{n_1}(1-X_1)^{n_2}X_2^{n_1+n_2}\right]^h, \end{aligned} \tag{8}$$

其中 X_1 和 X_2 是独立的, 且分别服从分布 $\beta(x|n_1-1, n_2-1)$ 和 $\beta(x|n_1+n_2-2, 1)$. 则 $\Pr\{V_1 \leqslant \nu\}$ 可以由积分得到. (参见习题 10.8 和习题 10.9.)

Anderson(1965a) 已经证明对于所有的 $\boldsymbol{a}$ 和置信系数 ε, $\boldsymbol{a}'\boldsymbol{\Sigma}_1\boldsymbol{a}/\boldsymbol{a}'\boldsymbol{\Sigma}_2\boldsymbol{a}$ 的置信区间由 $(l_p/U, l_1/L)$ 给出, 其中 $\Pr\{(n_2-p+1)L \leqslant n_2F_{n_1,n_2-p+1}\}\Pr\{(n_1-p+1)F_{n_1-p+1,n_2} \leqslant n_1U\} = 1-\varepsilon$.

10.6.2 方差分量

在 8.8 节中我们考虑过有固定效应的单因素方差分析的等价形式. 我们可以把平衡情况下 $(N_1 = N_2 = \cdots = N_q)$ 的模型记为

$$\begin{aligned} \boldsymbol{X}_\alpha^{(g)} &= \boldsymbol{\mu}^{(g)} + \boldsymbol{U}_\alpha^{(g)} \\ &= \boldsymbol{\mu} + \boldsymbol{\nu}_g + \boldsymbol{U}_\alpha^{(g)}, \quad \alpha = 1, \cdots, M, \quad g = 1, \cdots, q, \end{aligned} \tag{9}$$

其中 $\mathrm{E}(\boldsymbol{U}^{(g)}) = \boldsymbol{0}$, $\mathrm{E}(\boldsymbol{U}^{(g)}\boldsymbol{U}^{(g)\prime}) = \boldsymbol{\Sigma}$, $\boldsymbol{\nu}_g = \boldsymbol{\mu}^{(g)} - \boldsymbol{\mu}$, $\boldsymbol{\mu} = (1/q)\sum_{g=1}^q \boldsymbol{\mu}^{(g)}$ $(\sum_{g=1}^q \boldsymbol{\nu}_g = \boldsymbol{0})$. 无效应的原假设是 $\boldsymbol{\nu}_1 = \cdots = \boldsymbol{\nu}_q = \boldsymbol{0}$. 令 $\bar{\boldsymbol{x}}^{(g)} = (1/M)\sum_{\alpha=1}^M \boldsymbol{x}_\alpha^{(g)}$ 且 $\bar{\boldsymbol{x}} = (1/q)\sum_{g=1}^q \bar{\boldsymbol{x}}^{(g)}$. 方差分析表是

来源	平方和	自由度
效应	$\boldsymbol{H}=M\sum_{g=1}^{q}(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'$	$q-1$
误差	$\boldsymbol{G}=\sum_{g=1}^{q}\sum_{\alpha=1}^{M}(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}}^{(g)})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}}^{(g)})'$	$q(M-1)$
总和	$\sum_{g=1}^{q}\sum_{\alpha=1}^{M}(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'$	$qM-1$

无效应的原假设的不变检验是基于 $|\boldsymbol{H}-m\boldsymbol{G}|=0$ 或 $|\boldsymbol{S}_h-l\boldsymbol{S}_e|=0$ 的根的，其中 $\boldsymbol{S}_h=[1/(q-1)]\boldsymbol{H}$, $\boldsymbol{S}_e=[1/q(M-1)]\boldsymbol{G}$. 如果有一个或多个根太大则拒绝原假设. 误差矩阵 $\boldsymbol{G}$ 有分布 $W(\boldsymbol{\Sigma},q(M-1))$. 当原假设为真时效应矩阵有分布 $W(\boldsymbol{\Sigma},q-1)$, 当原假设不为真时它有非中心 Wishart 分布; 它的期望值是

$$\mathrm{E}(\boldsymbol{H})=(q-1)\boldsymbol{\Sigma}+M\sum_{g=1}^{q}(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu})(\boldsymbol{\mu}^{(g)}-\boldsymbol{\mu})' \tag{10}$$
$$=(q-1)\boldsymbol{\Sigma}+M\sum_{g=1}^{q}\boldsymbol{\nu}_g\boldsymbol{\nu}_g'.$$

带有随机效应的 MANOVA 模型是

$$\boldsymbol{X}_\alpha^{(g)}=\boldsymbol{\mu}+\boldsymbol{V}_g+\boldsymbol{U}_\alpha^{(g)},\quad \alpha=1,\cdots,M,\quad g=1,\cdots,q, \tag{11}$$

其中 $\boldsymbol{V}_g$ 有分布 $N(\boldsymbol{0},\boldsymbol{\Theta})$. 则 $\boldsymbol{X}_\alpha^{(g)}$ 有分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma}+\boldsymbol{\Theta})$. 无效应的原假设是

$$\boldsymbol{\Theta}=\boldsymbol{0}. \tag{12}$$

在这个模型中 $\boldsymbol{G}$ 也有分布 $W(\boldsymbol{\Sigma},q(M-1))$. 由于 $\bar{\boldsymbol{X}}^{(g)}=\boldsymbol{\mu}+\boldsymbol{V}_g+\bar{\boldsymbol{U}}^{(g)}$ 有分布 $N(\boldsymbol{\mu},(1/M)\boldsymbol{\Sigma}+\boldsymbol{\Theta})$, 所以 $\boldsymbol{H}$ 有分布 $W(\boldsymbol{\Sigma}+M\boldsymbol{\Theta},q-1)$. 原假设 (12) 等价于在这两个 Wishart 分布中协方差阵相等, 即 $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}+M\boldsymbol{\Theta}$. 矩阵 $\boldsymbol{G}$ 和 $\boldsymbol{H}$ 对应于 10.6.1 节中的 A_1 和 A_2. 但是, 这里原假设的另一种形式是 $(\boldsymbol{\Sigma}+M\boldsymbol{\Theta})-\boldsymbol{\Sigma}$ 是半正定的, 而不是 $\boldsymbol{\Sigma}\neq\boldsymbol{\Sigma}$. 如果 $\boldsymbol{H}$ 相对于 $\boldsymbol{G}$ 太大则拒绝原假设. 10.2 节中出现的任一准则都可以用来检验这里的原假设, 并且在原假设下它的分布和那里给出的一样.

检验 $\boldsymbol{\Theta}=\boldsymbol{0}$ 的似然比准则必须要把 $\boldsymbol{\Theta}$ 是半正定的事实考虑进去, 即在 Ω 下 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Sigma}+M\boldsymbol{\Theta}$ 的极大似然估计一定使得 $\boldsymbol{\Theta}$ 的估计为半正定的. 令 $l_1>l_2>\cdots>l_p$ 是

$$\left|\boldsymbol{H}-l\frac{1}{M-1}\boldsymbol{G}\right|=0 \tag{13}$$

的根. (注意到 $\{1/[q(M-1)]\}\boldsymbol{G}$ 和 $(1/q)\boldsymbol{H}$ 在不考虑 $\boldsymbol{\Theta}$ 为正定的情况下极大化似然函数.) 如果 $l_i>1$, 令 $l_1^*=l_i$, 如果 $l_i\leqslant 1$, 则令 $l_1^*=1$. 则检验原假设 $\boldsymbol{\Theta}=\boldsymbol{0}$ 相对于备择假设 $\boldsymbol{\Theta}$ 正定或 $\boldsymbol{\Theta}\neq\boldsymbol{0}$ 的似然比准则是

$$M^{\frac{1}{2}qMp}\prod_{i=1}^{p}\frac{l_i^{*\frac{1}{2}q}}{(l_i^*+M-1)^{\frac{1}{2}qM}}=M^{\frac{1}{2}qMk}\prod_{i=1}^{k}\frac{l_i^{\frac{1}{2}q}}{(l_i+M-1)^{\frac{1}{2}qM}}, \tag{14}$$

其中 k 是 (13) 式中大于 1 的根的个数. [参见 Anderson(1946b), (1984a), (1989a), Morris and Olkin(1964) 与 Klotz and Putter(1969).]

10.7 检验协方差阵与给定矩阵成正比的假设; 球形检验

10.7.1 假设

在很多单变量的统计分析中, 我们经常假定一个随机变量集是独立的并且有相同的方差. 在本节中我们考虑基于重复观测集的这些假定的检验.

更具体地, 我们用从 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 中抽出的 p 元向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 的样本来检验假设 $H: \boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, 其中 σ^2 不是特定的. 这个假设可以由 $\boldsymbol{\Sigma}$ 的特征根即

$$|\boldsymbol{\Sigma} - \phi \boldsymbol{I}| = 0 \tag{1}$$

的根给出一个代数解释. 这个假设是真的当且仅当 (1) 式的所有根都相等①. 表达它的另外一个方法是根 $\phi_1, \cdots, \phi_p$ 的算术均值等于其几何均值, 即

$$\frac{\prod_{i=1}^{p} \phi_i^{1/p}}{\sum_{i=1}^{p} \phi_i / p} = \frac{|\boldsymbol{\Sigma}|^{1/p}}{\operatorname{tr} \boldsymbol{\Sigma} / p} = 1. \tag{2}$$

有常数密度的椭球面的主轴的长度的平方与根 ϕ_i 成正比 (见第 11 章); 这个假设指出它们是相等的, 也就是, 椭球面是球面.

有了从 $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$ 中得到的观测向量 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ 后, 假设 H 等价于更一般的形式 $\boldsymbol{\Psi} = \sigma^2 \boldsymbol{\Psi}_0$, 其中 $\boldsymbol{\Psi}_0$ 是给定的. 令 $\boldsymbol{C}$ 是一个矩阵, 它满足

$$\boldsymbol{C} \boldsymbol{\Psi}_0 \boldsymbol{C}' = \boldsymbol{I}, \tag{3}$$

令 $\boldsymbol{\mu}^* = \boldsymbol{C}\boldsymbol{\nu}$, $\boldsymbol{\Sigma}^* = \boldsymbol{C}\boldsymbol{\Psi}\boldsymbol{C}'$, $\boldsymbol{x}_\alpha^* = \boldsymbol{C}\boldsymbol{y}_\alpha$. 则 $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_N^*$ 是来自 $N(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ 的观测, 且假设变为 $H: \boldsymbol{\Sigma}^* = \sigma^2 \boldsymbol{I}$.

10.7.2 准则

在典范型下, 假设 H 是假设 H_1 和假设 H_2 的组合, 其中 H_1 : $\boldsymbol{\Sigma}$ 是对角的或 $\boldsymbol{X}$ 的分量是独立的, H_2: 给定 $\boldsymbol{\Sigma}$ 是对角的条件下 $\boldsymbol{\Sigma}$ 的对角元是相等的或给定 $\boldsymbol{X}$ 的分量是独立的条件下 $\boldsymbol{X}$ 的分量的方差是相等的. 所以由引理 10.3.1 知, 对于 H 的似然比准则 λ 是对于 H_1 的准则 λ_1 和对于 H_2 的准则 λ_2 的乘积. 由 9.2 节我们可以看出, 对于 H_1 的准则是

$$\lambda_1 = \frac{|\boldsymbol{A}|^{\frac{1}{2}N}}{\prod a_{ii}^{\frac{1}{2}N}} = |r_{ij}|^{\frac{1}{2}N}, \tag{4}$$

其中

$$\boldsymbol{A} = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = (a_{ij}) \tag{5}$$

① 这可由 $\boldsymbol{\Sigma} = \boldsymbol{O}' \boldsymbol{\Phi} \boldsymbol{O}$ 得到, 其中 $\boldsymbol{\Phi}$ 是以根为对角元素的对角矩阵, $\boldsymbol{O}$是正交矩阵.

且 $r_{ij}=a_{ij}/\sqrt{a_{ii}a_{jj}}$. 把 $\boldsymbol{x}_\alpha$ 的第 i 个分量看作从第 i 个总体中抽出的第 α 个观测, 利用 10.2 节的结果我们可以得到 λ_2. (这里的 p 是 10.2 节中的 q, 这里的 N 是那里的 N_g, 这里的 pN 是那里的 N.) 因此

$$\begin{aligned}\lambda_2=&\frac{\prod_i\left[\sum_\alpha(x_{i\alpha}-\bar{x}_i)^2\right]^{\frac{1}{2}N}}{\left[\sum_{i,\alpha}(x_{i\alpha}-\bar{x}_i)^2/p\right]^{\frac{1}{2}pN}}\\=&\frac{\prod a_{ii}^{\frac{1}{2}N}}{(\operatorname{tr}\boldsymbol{A}/p)^{\frac{1}{2}pN}}.\end{aligned}\tag{6}$$

因此对于 H 的准则是

$$\lambda=\lambda_1\lambda_2=\frac{|\boldsymbol{A}|^{\frac{1}{2}N}}{(\operatorname{tr}\boldsymbol{A}/p)^{\frac{1}{2}pN}}.\tag{7}$$

可以看出 λ 和 (2) 式相似. 如果 $l_1,\cdots,l_p$ 是

$$|\boldsymbol{S}-l\boldsymbol{I}|=0\tag{8}$$

的根, 其中 $\boldsymbol{S}=(1/n)\boldsymbol{A}$, 则这个准则是几何平均和算术平均的比的幂,

$$\lambda=\left(\frac{\prod l_i^{1/p}}{\sum l_i/p}\right)^{\frac{1}{2}pN}.\tag{9}$$

给定从 $N(\boldsymbol{\nu},\boldsymbol{\Psi})$ 中得到的观测向量 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N$, 现在让我们回到假设 $\boldsymbol{\Psi}=\sigma^2\boldsymbol{\Psi}_0$. 用变换变量 $\{\boldsymbol{x}_\alpha\}^*$ 表示的准则是 $|\boldsymbol{A}^*|^{\frac{1}{2}N}(\operatorname{tr}\boldsymbol{A}^*/p)^{-\frac{1}{2}pN}$, 其中

$$\begin{aligned}\boldsymbol{A}^*=&\sum_{\alpha=1}^N(\boldsymbol{x}_\alpha^*-\bar{\boldsymbol{x}}^*)(\boldsymbol{x}_\alpha^*-\bar{\boldsymbol{x}}^*)'\\=&\boldsymbol{C}\sum_{\alpha=1}^N(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})'\boldsymbol{C}'\\=&\boldsymbol{CBC}',\end{aligned}\tag{10}$$

其中

$$\boldsymbol{B}=\sum_{\alpha=1}^N(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha-\bar{\boldsymbol{y}})'.\tag{11}$$

从 (3) 式中我们有 $\boldsymbol{\Psi}_0=\boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1}=(\boldsymbol{C}'\boldsymbol{C})^{-1}$. 因此

$$\begin{aligned}|\boldsymbol{A}^*|=&\frac{|\boldsymbol{B}|}{|\boldsymbol{\Psi}_0|}=|\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}|,\\\operatorname{tr}\boldsymbol{A}^*=&\operatorname{tr}\boldsymbol{CBC}'=\operatorname{tr}\boldsymbol{BC}'\boldsymbol{C}\\=&\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}.\end{aligned}\tag{12}$$

这些结果可以总结如下.

定理 10.7.1 给定从 $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$ 中得到的 p 元观测向量集 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$, 检验假设 $H: \boldsymbol{\Psi} = \sigma^2\boldsymbol{\Psi}_0$ 的似然比准则是

$$\frac{|\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}|^{\frac{1}{2}N}}{(\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/p)^{\frac{1}{2}pN}}, \tag{13}$$

其中 $\boldsymbol{\Psi}_0$ 是给定的而 σ^2 不是给定的.

Mauchly(1940) 给出了这个准则和在原假设下它的矩.

在原假设下 σ^2 的极大似然估计是 $\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/(pN)$, 这正是典范型的 $\operatorname{tr}\boldsymbol{A}/(pN)$; 一个无偏估计是 $\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/[p(N-1)]$, 或者典范型下的 $\operatorname{tr}\boldsymbol{A}/[p(N-1)]$ [Hotelling (1951)]. 则 $\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}/\sigma^2$ 服从自由度为 $p(N-1)$ 的 χ^2 分布.

10.7.3 准则的分布和矩

在原假设下似然比准则的分布可以由事实 $\lambda = \lambda_1\lambda_2$ 来刻画, 并且 λ_1 和 λ_2 是独立的, 故可由 λ_1 和 λ_2 的分布来确定. 正如我们在 7.6 节中看到的, 当 $\boldsymbol{\Sigma}$ 是对角矩阵时相关系数 $\{r_{ij}\}$ 的分布是与方差 $\{a_{ii}/(N-1)\}$ 独立的. 由于 λ_1 只依赖于 $\{r_{ij}\}$ 而 λ_2 只依赖于 $\{a_{ii}\}$, 所以当原假设为真时他们是独立分布的. 令 $W = \lambda^{2/N}$, $W_1 = \lambda_1^{2/N}$, $W_1 = \lambda_2^{2/N}$. 从定理 9.3.3 我们看到 W_1 有分布 $\prod_{i=2}^p X_i$, 其中 $X_2, \cdots, X_p$ 是独立的且 X_i 有密度 $\beta[x|\,\frac{1}{2}(n-i+1), \frac{1}{2}(i-1)]$, 其中 $n = N-1$. 从定理 10.4.2 和 $W_2 = p^pV_1^{2/N}$, 我们得到 W_2 服从分布 $p^p\prod_{j=2}^p Y_j^{j-1}(1-Y_j)$, 其中 $Y_2, \cdots, Y_p$ 是独立的且 Y_j 有密度 $\beta(y|\frac{1}{2}n(j-1), \frac{1}{2}n)$. 则 W 与 W_1W_2 同分布, 其中 W_1 和 W_2 是独立的.

W 的矩可以由这个特征或者定理 9.3.4 和定理 10.4.4 得到. 我们有

$$\mathrm{E}(W_1^h) = \frac{\Gamma^p\left(\frac{1}{2}n\right)}{\Gamma^p\left(\frac{1}{2}n+h\right)}\frac{\Gamma_p\left(\frac{1}{2}n+h\right)}{\Gamma_p\left(\frac{1}{2}n\right)}, \tag{14}$$

$$\mathrm{E}(W_2^h) = p^{hp}\frac{\Gamma^p\left(\frac{1}{2}n+h\right)\Gamma\left(\frac{1}{2}pn\right)}{\Gamma^p\left(\frac{1}{2}n\right)\Gamma\left(\frac{1}{2}pn+ph\right)}. \tag{15}$$

从而

$$\mathrm{E}(W^h) = p^{hp}\frac{\Gamma\left(\frac{1}{2}pn\right)}{\Gamma\left(\frac{1}{2}pn+ph\right)}\frac{\Gamma_p\left(\frac{1}{2}n+h\right)}{\Gamma_p\left(\frac{1}{2}n\right)}. \tag{16}$$

对于 $p=2$, 利用伽玛函数的倍量公式我们有

$$\begin{aligned}\mathrm{E}(W^h) &= 4^h\frac{\Gamma(n)}{\Gamma(n+2h)}\prod_{i=1}^2\frac{\Gamma\left[\frac{1}{2}(n+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n+1-i)\right]} \\ &= \frac{\Gamma(n)\Gamma(n-1+2h)}{\Gamma(n+2h)\Gamma(n-1)} = \frac{n-1}{n-1+2h} \\ &= (n-1)\int_0^1 z^{n-2+2h}\mathrm{d}z.\end{aligned} \tag{17}$$

因此 W 与 Z^2 同分布, 其中 Z 有密度 $(n-1)\,z^{n-2}$, 故 W 有密度 $\frac{1}{2}(n-1)w^{\frac{1}{2}(n-3)}$. 其分布函数为

$$\Pr\{W\leqslant w\}=F(w)=w^{\frac{1}{2}(n-1)}. \tag{18}$$

这个结果也可以由 (8) 式的根 l_1,l_2 的联合分布得到. $p=3,4,6$ 时的密度已经由 Consul(1967b) 得到. 也可参见 Pillai and Nagarsenkar(1971).

10.7.4 分布的渐近展开

由 (16) 式我们看到 $W^{\frac{1}{2}n}=Z$ 的 r 阶矩为

$$\mathrm{E}(Z^r)=Kp^{\frac{1}{2}npr}\frac{\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}n(1+r)+\frac{1}{2}(1-i)\right]}{\Gamma\left[\frac{1}{2}pn(1+r)\right]}. \tag{19}$$

这和 8.5 节中 (1) 式的形式相同, 其中

$$\begin{aligned}&a=p, && x_k=\tfrac{1}{2}n, && \xi_k=\tfrac{1}{2}(1-k), && k=1,\cdots,p,\\ &b=1, && y_1=\tfrac{1}{2}np, && \eta_1=0.\end{aligned} \tag{20}$$

因此 8.5 节的展开是有效的, 其中 $f=\frac{1}{2}p(p+1)-1$. 为了使展开式中第二项为零, 我们取 ρ 满足

$$1-\rho=\frac{2p^2+p+2}{6pn}. \tag{21}$$

则

$$\omega_2=\frac{(p+2)(p-1)(p-2)(2p^3+6p^2+3p+2)}{288p^2n^2\rho^2}. \tag{22}$$

因此 W 的分布函数可以由下式得到

$$\begin{aligned}&\Pr\{-2\rho\ln Z\leqslant z\}\\ =&\Pr\{-n\rho\ln W\leqslant z\}\\ =&\Pr\left\{\chi_f^2\leqslant z\right\}+\omega_2\left(\Pr\left\{\chi_{f+4}^2\leqslant z\right\}-\Pr\left\{\chi_f^2\leqslant z\right\}\right)+O(n^{-3}).\end{aligned} \tag{23}$$

因子 $c(n,p,\varepsilon)$ 已经在表 B.6 中给出, 它使得

$$\Pr\left\{-n\rho\ln W\leqslant c(n,p,\varepsilon)\,\chi^2_{\frac{1}{2}p(n+1)-1}(\varepsilon)\right\}=\varepsilon. \tag{24}$$

Nagarsenkar and Pillai(1973a) 有关于 W 的表.

10.7.5 不变检验

原假设 $H:\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$ 关于变换 $\boldsymbol{X}^*=c\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{\nu}$ 是不变的, 其中 c 是一个标量, $\boldsymbol{Q}$ 是一个正交矩阵. 在位置变换下充分统计量的不变量是 $\boldsymbol{A}$, 在正交变换下 $\boldsymbol{A}$ 的不变量是特征根 $l_1,\cdots,l_p$, 在标度变换下根的不变量是齐次度为 0 的函数, 例如根的比 $l_1/l_2,\cdots,l_{p-1}/l_p$. 不变检验是基于这样的函数的, 似然比准则就是这样的一个函数.

Nagao(1973a) 提出准则

$$\begin{aligned}
&\frac{1}{2}n\,\mathrm{tr}\left(\boldsymbol{S}-\frac{\mathrm{tr}\,\boldsymbol{S}}{p}\boldsymbol{I}\right)\frac{p}{\mathrm{tr}\,\boldsymbol{S}}\left(\boldsymbol{S}-\frac{\mathrm{tr}\,\boldsymbol{S}}{p}\boldsymbol{I}\right)\frac{p}{\mathrm{tr}\,\boldsymbol{S}} \qquad (25)\\
=&\frac{1}{2}n\,\mathrm{tr}\left(\frac{p}{\mathrm{tr}\,\boldsymbol{S}}\boldsymbol{S}-\boldsymbol{I}\right)^2=\frac{1}{2}n\left[\frac{p^2}{(\mathrm{tr}\,\boldsymbol{S})^2}\mathrm{tr}\,\boldsymbol{S}^2-p\right]\\
=&\frac{1}{2}n\left[\frac{p^2}{(\sum_{i=1}^p l_i)^2}\sum_{i=1}^p l_i^2-p\right]=\frac{1}{2}n\frac{\sum_{i=1}^p(l_i-\bar{l})^2}{\bar{l}^2},
\end{aligned}$$

其中 $\bar{l}=\sum_{i=1}^p l_i/p$. (25) 式的左边基于 7.8 节的损失函数 $L_p(\boldsymbol{\Sigma},\boldsymbol{G})$, 右边表明它和样本协方差阵 $\boldsymbol{S}$ 的特征根的变异系数的平方是成比例的. 另一个准则是 l_1/l_p. 百分位点已经由 Krishnaiah and Schuurmann(1974) 给出.

10.7.6 置信区域

给定 $N(\boldsymbol{\nu},\boldsymbol{\Psi})$ 中的观测 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N$, 对于任意给定的 $\boldsymbol{\Psi}_0$, 我们可以检验 $\boldsymbol{\Psi}=\sigma^2\boldsymbol{\Psi}_0$. 从这个检验类中我们可以建立 $\boldsymbol{\Psi}$ 的一个置信区域. 如果矩阵在置信区域中, 则它的所有幂都在里面. 如果 $\boldsymbol{y}_\alpha$ 的所有分量在同一个单位圆中, 但是研究者想要一个与这个公共圆独立的区域, 这类置信区域是我们感兴趣的. 置信度为 $1-\varepsilon$ 的置信区域包含所有满足下面条件的矩阵 $\boldsymbol{\Psi}^*$,

$$\frac{|\boldsymbol{B}\boldsymbol{\Psi}^{*-1}|}{\left[(\mathrm{tr}\,\boldsymbol{B}\boldsymbol{\Psi}^{*-1})/p\right]^p}\geqslant\lambda^{2/N}(\varepsilon), \qquad (26)$$

其中 $\lambda(\varepsilon)$ 是这个准则的 ε 显著性水平.

考虑 $p=2$ 的情形. 如果这个度量的公共圆是不相关的, 研究者对 $\tau=\psi_{11}/\psi_{22}$ 和 $\rho=\psi_{12}/\sqrt{\psi_{11}\psi_{22}}$ 感兴趣. 在这种情况下

$$\begin{aligned}
\boldsymbol{\Psi}^{-1}&=\frac{1}{\psi_{11}\psi_{22}(1-\rho^2)}\begin{pmatrix}\psi_{22} & -\rho\sqrt{\psi_{11}\psi_{22}}\\ -\rho\sqrt{\psi_{11}\psi_{22}} & \psi_{11}\end{pmatrix} \qquad (27)\\
&=\frac{1}{\psi_{11}(1-\rho^2)}\begin{pmatrix}1 & -\rho\sqrt{\tau}\\ -\rho\sqrt{\tau} & \tau\end{pmatrix}.
\end{aligned}$$

由 τ 和 ρ 表示的区域是

$$4\frac{(b_{11}b_{22}-b_{12}^2)(1-\rho^2)^\tau}{(b_{11}+\tau b_{22}-2\rho\sqrt{\tau}b_{12})^2}\geqslant\lambda^{2/N}(\varepsilon). \qquad (28)$$

Hickman(1953) 给出了这样一个置信区域的例子.

10.8 检验一个协方差阵等于一个给定的矩阵的假设

10.8.1 准则

如果 Y 服从分布 $N(\boldsymbol{\nu},\boldsymbol{\Psi})$, 我们想要检验 $H_1:\boldsymbol{\Psi}=\boldsymbol{\Psi}_0$, 其中 $\boldsymbol{\Psi}_0$ 是一个给定

的正定矩阵. 从上一节的讨论中我们知道这等价于检验假设 $H_1: \boldsymbol{\Sigma} = \boldsymbol{I}$, 其中 $\boldsymbol{\Sigma}$ 是服从分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的向量 X 的协方差阵. 给定一个样本 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$, 似然比准则是

$$\lambda_1 = \frac{\max_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \boldsymbol{I})}{\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}, \tag{1}$$

其中似然函数是

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right]. \tag{2}$$

第 3 章的结果说明

$$\begin{aligned}\lambda_1 &= \frac{(2\pi)^{-\frac{1}{2}pN}\exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})\right]}{(2\pi)^{-\frac{1}{2}pN}|(1/N)\,\boldsymbol{A}|^{-\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}pN}} \\ &= \left(\frac{\mathrm{e}}{N}\right)^{\frac{1}{2}pN}|\boldsymbol{A}|^{\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}},\end{aligned} \tag{3}$$

其中

$$\boldsymbol{A} = \sum_{\alpha}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'. \tag{4}$$

Sugiura and Nagao(1968) 已经证明这个似然比检验是有偏的, 但是修正后的基于下式的似然比准则是无偏的,

$$\lambda_1^* = \left(\frac{\mathrm{e}}{n}\right)^{\frac{1}{2}pn}|\boldsymbol{A}|^{\frac{1}{2}n}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}} = \mathrm{e}^{\frac{1}{2}pn}(|\boldsymbol{S}|\mathrm{e}^{-\operatorname{tr}\boldsymbol{S}})^{\frac{1}{2}n} \tag{5}$$

其中 $\boldsymbol{S} = (1/n)\boldsymbol{A}$. 需要注意的是

$$-\frac{2}{n}\ln\lambda_1^* = \operatorname{tr}\boldsymbol{S} - \ln|\boldsymbol{S}| - p = L_l(\boldsymbol{I}, \boldsymbol{S}), \tag{6}$$

其中 $L_l(\boldsymbol{I}, \boldsymbol{S})$ 是用 7.8 节中 (2) 式定义的 $\boldsymbol{S}$ 来估计 $\boldsymbol{I}$ 的损失函数. 由 $\boldsymbol{S}$ 的特征根表示的准则 (6) 是一个常数加上

$$\sum_{i=1}^{p} l_i - \ln\prod_{i=1}^{p} l_i - p = \sum_{i=1}^{p}(l_i - \ln l_i - 1). \tag{7}$$

对于每个 i, (7) 式在 $l_i = 1$ 取极小值.

利用上一节的代数知识, 我们知道给定 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ 作为 $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$ 中的 p 元观测向量, 检验假设 $H_1: \boldsymbol{\Psi} = \boldsymbol{\Psi}_0$ 的修正的似然比准则是

$$\lambda_1^* = \left(\frac{\mathrm{e}}{n}\right)^{\frac{1}{2}pn}|\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}|^{\frac{1}{2}n}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}}, \tag{8}$$

其中 $\boldsymbol{\Psi}_0$ 是已知的, 且

$$\boldsymbol{B} = \sum_{\alpha=1}^{N}(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})(\boldsymbol{y}_\alpha - \bar{\boldsymbol{y}})'. \tag{9}$$

10.8.2 修正的似然比准则的分布和矩

原假设 $H_1:\boldsymbol{\Sigma}=\boldsymbol{I}$ 是 10.7 节的原假设 $H:\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$ 和给定 $\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$ 下原假设 $\sigma^2=1$ 的交集. 由 (3) 式给出的 H_1 的似然比准则是 10.7 节的 (7) 式和

$$\left(\frac{\operatorname{tr}\boldsymbol{A}}{pN}\right)^{\frac{1}{2}pN}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}+\frac{1}{2}pN} \tag{10}$$

的乘积, 上式也是给定 $H:\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$ 下检验假设 $\sigma^2=1$ 的似然比准则. 修正的准则 λ_1^* 是 $|\boldsymbol{A}|^{\frac{1}{2}n}/(\operatorname{tr}\boldsymbol{A}/p)^{\frac{1}{2}pn}$ 和

$$\left(\frac{\operatorname{tr}\boldsymbol{A}}{pn}\right)^{\frac{1}{2}pn}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}+\frac{1}{2}pn} \tag{11}$$

的乘积, 这两个因子是独立的 (引理 10.4.1). 修正的准则的分布特征可以由 10.7.3 节得到. 在原假设下 $\operatorname{tr}\boldsymbol{A}$ 服从自由度为 np 的 χ^2 分布.

我们可以利用 $\boldsymbol{A}$ 服从分布 $W(\boldsymbol{\Sigma},n)$ 的事实来得到 λ_1^*[由 (5) 式定义] 的矩和特征函数, 而不是由上面的特征得到. 我们下面计算

$$\begin{aligned}\mathrm{E}(\lambda_1^{*h})&=\int\cdots\int\left(\frac{\mathrm{e}^{\frac{1}{2}pn}}{n^{\frac{1}{2}pn}}|\boldsymbol{A}|^{\frac{1}{2}n}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{A}}\right)^h w(\boldsymbol{A}|\boldsymbol{\Sigma},n)\mathrm{d}\boldsymbol{A} \\ &=\frac{\mathrm{e}^{\frac{1}{2}pnh}}{n^{\frac{1}{2}pnh}}\int\cdots\int|\boldsymbol{A}|^{\frac{1}{2}nh}\mathrm{e}^{-\frac{1}{2}h\operatorname{tr}\boldsymbol{A}}w(\boldsymbol{A}|\boldsymbol{\Sigma},n)\mathrm{d}\boldsymbol{A}.\end{aligned} \tag{12}$$

因为

$$\begin{aligned}|\boldsymbol{A}|^{\frac{1}{2}nh}\mathrm{e}^{-\frac{1}{2}h\operatorname{tr}\boldsymbol{A}}w(\boldsymbol{A}|\boldsymbol{\Sigma},n)&=\frac{|\boldsymbol{A}|^{\frac{1}{2}(n+nh-p-1)}\mathrm{e}^{-\frac{1}{2}(\operatorname{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{A}+\operatorname{tr}h\boldsymbol{A})}}{2^{\frac{1}{2}pn}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\Gamma_p\left(\frac{1}{2}n\right)} \\ &=\frac{2^{\frac{1}{2}pnh}\Gamma_p\left[\frac{1}{2}n(1+h)\right]}{|\boldsymbol{\Sigma}^{-1}+h\boldsymbol{I}|^{\frac{1}{2}(n+nh)}|\boldsymbol{\Sigma}|^{\frac{1}{2}n}\Gamma_p\left(\frac{1}{2}n\right)} \\ &\quad\cdot\frac{|\boldsymbol{\Sigma}^{-1}+h\boldsymbol{I}|^{\frac{1}{2}(n+nh)}|\boldsymbol{A}|^{\frac{1}{2}(n+nh-p-1)}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}+h\boldsymbol{I})\boldsymbol{A}}}{2^{\frac{1}{2}p(n+nh)}\Gamma_p\left[\frac{1}{2}n(1+h)\right]} \\ &=\frac{2^{\frac{1}{2}pnh}|\boldsymbol{\Sigma}|^{\frac{1}{2}nh}\Gamma_p\left[\frac{1}{2}n(1+h)\right]}{|\boldsymbol{I}+h\boldsymbol{\Sigma}|^{\frac{1}{2}(n+nh)}\Gamma_p\left(\frac{1}{2}n\right)} \\ &\quad\cdot w\left(\boldsymbol{A}|(\boldsymbol{\Sigma}^{-1}+h\boldsymbol{I})^{-1},n+nh\right),\end{aligned} \tag{13}$$

λ_1^* 的 h 阶矩为

$$\mathrm{E}\lambda_1^{*h}=\left(\frac{2\mathrm{e}}{n}\right)^{\frac{1}{2}phn}\frac{|\boldsymbol{\Sigma}|^{\frac{1}{2}nh}\prod_{j=1}^p\Gamma\left[\frac{1}{2}(n+nh+1-j)\right]}{|\boldsymbol{I}+h\boldsymbol{\Sigma}|^{\frac{1}{2}(n+nh)}\prod_{j=1}^p\Gamma\left[\frac{1}{2}(n+1-j)\right]}. \tag{14}$$

则 $-2\ln\lambda_1^*$ 的特征函数是

$$\begin{aligned}\mathrm{E}(\mathrm{e}^{-2it\ln\lambda_1^*}) &= \mathrm{E}(\lambda_1^{*-2it}) \\ &= \left(\frac{2\mathrm{e}}{n}\right)^{-ipnt} \frac{|\boldsymbol{\Sigma}|^{-int}}{|\boldsymbol{I}-2\mathrm{i}t\boldsymbol{\Sigma}|^{\frac{1}{2}n-int}} \cdot \prod_{j=1}^{p} \frac{\Gamma\left[\frac{1}{2}(n+1-j)-\mathrm{i}nt\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}.\end{aligned} \tag{15}$$

当原假设为真时, $\boldsymbol{\Sigma}=\boldsymbol{I}$, 且

$$\mathrm{E}(\mathrm{e}^{-2it\ln\lambda_1^*}) = \left(\frac{2\mathrm{e}}{n}\right)^{-ipnt} (1-2\mathrm{i}t)^{-\frac{1}{2}p(n-2int)} \prod_{j=1}^{p} \frac{\Gamma\left[\frac{1}{2}(n+1-j)-\mathrm{i}nt\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}. \tag{16}$$

这个特征函数是如下 p 项的乘积

$$\phi_j(t) = \left(\frac{2\mathrm{e}}{n}\right)^{-int} ((1-2\mathrm{i}t)^{-\frac{1}{2}(n-2int)} \frac{\Gamma\left[\frac{1}{2}(n+1-j)-\mathrm{i}nt\right]}{\Gamma\left[\frac{1}{2}(n+1-j)\right]}. \tag{17}$$

因此 $-2\ln\lambda_1^*$ 服从 p 个独立变量的和的分布, 第 j 个的特征函数是 (17) 式. 利用伽玛函数的 Stirling 近似, 我们有

$$\begin{aligned}\phi_j(t) &\sim 2^{-\mathrm{i}nt}\mathrm{e}^{-\mathrm{i}nt}n^{\mathrm{i}nt}(1-2\mathrm{i}t)^{\frac{1}{2}(2int-n)} \\ &\quad \cdot \frac{\mathrm{e}^{-\left[\frac{1}{2}(n+1-j)-int\right]}\left[\frac{1}{2}(n+1-j)-\mathrm{i}nt\right]^{\frac{1}{2}(n-j)-\mathrm{i}nt}}{\mathrm{e}^{-\left[\frac{1}{2}(n+1-j)\right]}\left[\frac{1}{2}(n+1-j)\right]^{\frac{1}{2}(n-j)}} \\ &= (1-2\mathrm{i}t)^{-\frac{1}{2}j}\left(1-\frac{\mathrm{i}t(j-1)}{\frac{1}{2}(n-j+1)(1-2\mathrm{i}t)}\right)^{\frac{1}{2}(n+1-j)-\frac{1}{2}} \\ &\quad \cdot \left(1-\frac{2j-1}{n(1-2\mathrm{i}t)}\right)^{-int}.\end{aligned} \tag{18}$$

当 $n\to\infty$ 时, $\phi_j(t)\to(1-2\mathrm{i}t)^{-\frac{1}{2}j}$, 这正是 χ_j^2 (有 j 个自由度的 χ^2) 的特征函数. 因此 $-2\ln\lambda_1^*$ 是渐近地服从分布 $\sum_{j=1}^{p}\chi_j^2$ 的, 这是一个自由度为 $\sum_{j=1}^{p}j=\frac{1}{2}p(p+1)$ 的 χ^2. λ_1^* 的分布可以近一步展开为 [Korin(1968), Davis(1971)]

$$\begin{aligned}&\Pr\left\{-2\rho\ln\lambda_1^*\leqslant z\right\} \\ &= \Pr\left\{\chi_f^2\leqslant z\right\}+\frac{\gamma_2}{\rho^2N^2}\left(\Pr\left\{\chi_{f+4}^2\leqslant z\right\}-\Pr\left\{\chi_f^2\leqslant z\right\}\right)+O(N^{-3}),\end{aligned} \tag{19}$$

其中

$$\rho = 1-\frac{2p^2+3p-1}{6N(p+1)}, \tag{20}$$

$$\gamma_2 = \frac{p(2p^4+6p^3+p^2-12p-13)}{288(p+1)}. \tag{21}$$

对于 $p=2(1)10$ 和 $n=6(1)30(5)50, 60, 120$ 的情况, Nagarsenker and Pillai (1973b) 得到了精确分布并且给出了 5% 和 1% 的分位数表, 正如 Davis and Field(1971) 所做的一样. 表 B.7[出于 Korin(1968)] 给出了当 n 很小和 $p=2(1)10$ 时 $-2\ln\lambda_1^*$ 的 5% 和 1% 的分位数.

10.8.3 不变检验

原假设 $H:\boldsymbol{\Sigma}=\boldsymbol{I}$ 关于变换 $\boldsymbol{X}^*=\boldsymbol{QX}+\boldsymbol{\nu}$ 是不变的, 其中 Q 是正交矩阵. 充分统计量的不变量是 $\boldsymbol{S}$ 的特征根 $l_1,\cdots,l_p$, 参数的不变量是 $\boldsymbol{\Sigma}$ 的特征根. 不变检验是基于 $\boldsymbol{S}$ 的根的, 修正的似然比准则就是其中的一个. Nagao(1973a) 提出准则

$$\frac{1}{2}n\,\mathrm{tr}\,(\boldsymbol{S}-\boldsymbol{I})^2=\frac{1}{2}n\sum_{i=1}^{p}(l_i-1)^2. \tag{22}$$

在原假设下这个准则有自由度为 $\frac{1}{2}p(p+1)$ 的极限 χ^2 分布.

在 Roy (1957) 6.4 节中有一个基于最大特征根和最小特征根 l_1 和 l_p 的准则: 如果

$$l_p<l \quad 或者 \quad l_1>u, \tag{23}$$

则拒绝原假设, 其中

$$\Pr(l<l_p,l_1<u|\boldsymbol{\Sigma}=\boldsymbol{I})=1-\varepsilon \tag{24}$$

且 ε 是显著性水平. Clemin, Krishnaiah, and Waikar(1973) 给出了 $u=1/l$ 时的表. 也可参见 Schuurman and Waikar(1973).

10.8.4 二次型的置信界

基于最小特征根和最大特征根的检验方法可以转换成求 $\boldsymbol{\Sigma}$ 的二次型的置信界问题. 假设 $n\boldsymbol{S}$ 有分布 $W(\boldsymbol{\Sigma},n)$. 令 $\boldsymbol{C}$ 是一个非奇异矩阵且满足 $\boldsymbol{\Sigma}=\boldsymbol{C}'\boldsymbol{C}$. 则 $n\boldsymbol{S}^*=n\boldsymbol{C}'^{-1}\boldsymbol{S}\boldsymbol{C}^{-1}$ 有分布 $W(\boldsymbol{I},n)$. 因为对于所有的 $\boldsymbol{a}$ 有 $l_p^*\leqslant\boldsymbol{a}'\boldsymbol{S}^*\boldsymbol{a}/\boldsymbol{a}'\boldsymbol{a}<l_1^*$, 其中 l_p^* 和 l_1^* 是 $\boldsymbol{S}^*$ 的最小特征根和最大特征根 (11.2 节和 A.2 节), 所以

$$\Pr\left\{l\leqslant\frac{\boldsymbol{a}'\boldsymbol{S}^*\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}\leqslant u,\ \forall\boldsymbol{a}\neq\boldsymbol{0}\right\}=1-\varepsilon, \tag{25}$$

其中

$$\Pr\{l\leqslant l_p^*\leqslant l_1^*\leqslant u\}=1-\varepsilon. \tag{26}$$

令 $\boldsymbol{a}=\boldsymbol{Cb}$, 则 $\boldsymbol{a}'\boldsymbol{a}=\boldsymbol{b}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{b}=\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}$, $\boldsymbol{a}'\boldsymbol{S}^*\boldsymbol{a}=\boldsymbol{b}'\boldsymbol{C}'\boldsymbol{S}^*\boldsymbol{C}\boldsymbol{b}=\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}$. 因此 (25) 式为

$$\begin{aligned}1-\varepsilon&=\Pr\left\{l\leqslant\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}\leqslant u,\quad\forall\boldsymbol{b}\neq\boldsymbol{0}\right\}\\&=\Pr\left\{\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{u}\leqslant\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}\leqslant\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{l},\quad\forall\boldsymbol{b}\right\}.\end{aligned} \tag{27}$$

给定一个观测 $\boldsymbol{S}$, 我们可以以置信水平 $1-\varepsilon$ 断言

$$\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{u}\leqslant\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}\leqslant\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{l},\quad\forall\boldsymbol{b}. \tag{28}$$

如果 $\boldsymbol{b}$ 在第 i 个位置上为 1 而在其余位置为 0, 则 (28) 式变为 $s_{ii}/u\leqslant\sigma_{ii}\leqslant s_{ii}/l$. 如果 $\boldsymbol{b}$ 在第 i 个位置上为 1, 在第 j 个位置上为 -1, 且 $i\neq j$, 而在其余位置为 0, 则 (28) 式变为

$$\frac{s_{ii}+s_{jj}-2s_{ij}}{u} \leqslant \sigma_{ii}+\sigma_{jj}-2\sigma_{ij} \leqslant \frac{s_{ii}+s_{jj}-2s_{ij}}{l}. \tag{29}$$

对这些不等式进行运算得到

$$\frac{s_{ij}}{l}-\frac{s_{ii}+s_{jj}}{2}\left(\frac{1}{l}-\frac{1}{u}\right) \leqslant \sigma_{ij} \leqslant \frac{s_{ij}}{u}+\frac{s_{ii}+s_{jj}}{2}\left(\frac{1}{l}-\frac{1}{u}\right), \quad i \neq j. \tag{30}$$

我们可以同时得到 $\boldsymbol{\Sigma}$ 的所有元素的置信区间.

从 (27) 式我们可以得到

$$\begin{aligned} 1-\varepsilon &= \Pr\left\{\frac{1}{u}\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{b}} \leqslant \frac{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{b}} \leqslant \frac{1}{l}\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{b}}, \quad \forall \boldsymbol{b}\right\} \\ &\leqslant \Pr\left\{\frac{1}{u}\min_{\boldsymbol{a}}\frac{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}} \leqslant \frac{\boldsymbol{b}'\boldsymbol{\Sigma}\boldsymbol{b}}{\boldsymbol{b}'\boldsymbol{b}} \leqslant \frac{1}{l}\max_{\boldsymbol{a}}\frac{\boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{a}}, \quad \forall \boldsymbol{b}\right\} \\ &= \Pr\left\{\frac{1}{u}l_p \leqslant \lambda_p \leqslant \lambda_1 \leqslant \frac{1}{l}l_1\right\}, \end{aligned} \tag{31}$$

其中 l_1 和 l_p 是 $\boldsymbol{S}$ 的最大特征根和最小特征根而 λ_1 和 λ_p 是 $\boldsymbol{\Sigma}$ 的最大特征根和最小特征根. 则

$$\frac{1}{u}l_p \leqslant \lambda(\boldsymbol{\Sigma}) \leqslant \frac{1}{l}l_1 \tag{32}$$

是 $\boldsymbol{\Sigma}$ 的所有特征根的置信水平为 $1-\varepsilon$ 的置信区间. 在 11.6 节我们将给出 $\lambda(\boldsymbol{\Sigma})$ 在精确置信水平下的更紧凑的界.

10.9　检验均值向量和协方差阵分别等于给定的向量和矩阵的假设

在第 3 章中我们指出, 如果 $\boldsymbol{\Psi}$ 已知, $(\bar{\boldsymbol{y}}-\boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\nu}_0)$ 适合于检验

$$H_2: \boldsymbol{\nu}=\boldsymbol{\nu}_0, \quad \text{给定 } \boldsymbol{\Psi} \equiv \boldsymbol{\Psi}_0. \tag{1}$$

现在我们把 10.8 节中的 H_1 和与这里的 H_2 结合起来, 基于 $N(\boldsymbol{\nu},\boldsymbol{\Psi})$ 中的一组样本 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N$ 来检验

$$H: \boldsymbol{\nu}=\boldsymbol{\nu}_0, \quad \boldsymbol{\Psi}=\boldsymbol{\Psi}_0. \tag{2}$$

令

$$\boldsymbol{X}=\boldsymbol{C}(\boldsymbol{Y}-\boldsymbol{\nu}_0), \tag{3}$$

其中

$$\boldsymbol{C}\boldsymbol{\Psi}_0\boldsymbol{C}'=\boldsymbol{I}. \tag{4}$$

则 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 组成了从 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 中得到的样本, 并且假设是

$$H: \boldsymbol{\mu}=\boldsymbol{0}, \quad \boldsymbol{\Sigma}=\boldsymbol{I}. \tag{5}$$

给定 $\boldsymbol{\Sigma}=\boldsymbol{I}$, 检验 $H_2: \boldsymbol{\mu}=\boldsymbol{0}$ 的似然比准则是

$$\lambda_2 = \mathrm{e}^{-\frac{1}{2}N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}}. \tag{6}$$

检验 H 的似然比准则是 (由引理 10.3.1)

$$\begin{aligned}\lambda = \lambda_1\lambda_2 &= \left(\frac{\mathrm{e}}{N}\right)^{\frac{1}{2}pN} |\boldsymbol{A}|^{\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,\boldsymbol{A}}\mathrm{e}^{-\frac{1}{2}N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}}} \\ &= \left(\frac{\mathrm{e}}{N}\right)^{\frac{1}{2}pN} |\boldsymbol{A}|^{\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\,(\boldsymbol{A}+N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}')} \\ &= \left(\frac{\mathrm{e}}{N}\right)^{\frac{1}{2}pN} |\boldsymbol{A}|^{\frac{1}{2}N}\mathrm{e}^{-\frac{1}{2}\sum \boldsymbol{x}_\alpha'\boldsymbol{x}_\alpha}.\end{aligned} \tag{7}$$

似然比检验 (如果 λ 小于一个合适的常数则拒绝 H) 是无偏的 [Srivastava and Khatri(1979), 定理 10.4.5]. 两个因子 λ_1 和 λ_2 是独立的, 因为 λ_1 是 $\boldsymbol{A}$ 的函数而 λ_2 是 $\bar{\boldsymbol{x}}$ 的函数, 且 $\boldsymbol{A}$ 和 $\bar{\boldsymbol{x}}$ 是独立的. 因为

$$\mathrm{E}(\lambda_2^h) = \mathrm{E}\left(\mathrm{e}^{-\frac{1}{2}hN\sum \bar{x}_i^2}\right) = \mathrm{E}\left(\mathrm{e}^{-\frac{1}{2}h\chi_p^2}\right) = (1+h)^{-\frac{1}{2}p}, \tag{8}$$

在原假设下, λ 的 h 阶矩是

$$\mathrm{E}(\lambda^h) = \mathrm{E}(\lambda_1^h)\mathrm{E}(\lambda_2^h) = \left(\frac{2\mathrm{e}}{N}\right)^{\frac{1}{2}pNh} \frac{1}{(1+h)^{\frac{1}{2}pN(1+h)}} \frac{\Gamma_p\left[\frac{1}{2}(n+Nh)\right]}{\Gamma_p\left(\frac{1}{2}n\right)}. \tag{9}$$

则

$$-2\ln\lambda = -2\ln\lambda_1 - 2\ln\lambda_2 \tag{10}$$

渐近地服从自由度为 $f = p(p+1)/2 + p$ 的 χ^2 分布. 事实上, $-2\rho\ln\lambda$ 的分布 [Davis(1971)] 的渐近展开是

$$\begin{aligned}&\Pr\{-2\rho\ln\lambda \leqslant z\} \\ =&\Pr\{\chi_f^2 \leqslant z\} + \frac{\gamma_2}{\rho^2N^2}\left(\Pr\{\chi_{f+4}^2 \leqslant z\} - \Pr\{\chi_f^2 \leqslant z\}\right) + O(N^{-3}),\end{aligned} \tag{11}$$

其中

$$\rho = 1 - \frac{2p^2+9p-11}{6N(p+3)}, \tag{12}$$

$$\gamma_2 = \frac{p(2p^4+18p^3+49p^2+36p-13)}{288(p-3)}. \tag{13}$$

Nagarsenker and Pillai(1974) 利用矩得到了当 $p = 2(1)6$ 和 $N = 4(1)20(2)40(5)100$ 的精确分布, 并制出了 5% 和 1% 的分位数表.

现在让我们回到观测 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$. 则

$$\begin{aligned}\sum_\alpha \boldsymbol{x}_\alpha'\boldsymbol{x}_\alpha &= \sum_\alpha(\boldsymbol{y}_\alpha - \boldsymbol{\nu}_0)'\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{y}_\alpha - \boldsymbol{\nu}_0) \\ &= \sum_\alpha(\boldsymbol{y}_\alpha - \boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\boldsymbol{y}_\alpha - \boldsymbol{\nu}_0) \\ &= \mathrm{tr}\,\boldsymbol{A} + N\bar{\boldsymbol{x}}'\bar{\boldsymbol{x}} \\ &= \mathrm{tr}\,(\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}) + N(\bar{\boldsymbol{y}} - \boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\nu}_0),\end{aligned} \tag{14}$$

$$|\boldsymbol{A}| = |\boldsymbol{B\Psi}_0^{-1}|. \tag{15}$$

定理 10.9.1 给定 $N(\boldsymbol{\nu}, \boldsymbol{\Psi})$ 中的 p 元观测向量 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$, 检验假设 $H: \boldsymbol{\nu} = \boldsymbol{\nu}_0, \boldsymbol{\Psi} = \boldsymbol{\Psi}_0$ 的似然比准则是

$$\lambda = \left(\frac{\mathrm{e}}{N}\right)^{\frac{1}{2}pN} |\boldsymbol{B\Psi}_0^{-1}|^{\frac{1}{2}N} \mathrm{e}^{-\frac{1}{2}[\operatorname{tr} \boldsymbol{B\Psi}_0^{-1} + N(\bar{\boldsymbol{y}}-\boldsymbol{\nu}_0)'\boldsymbol{\Psi}_0^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\nu}_0)]}. \tag{16}$$

当原假设为真时, $-2\ln\lambda$ 渐近地服从自由度为 $\frac{1}{2}p(p+1)+p$ 的 χ^2 分布.

10.10 检验的容许性

我们将考虑对于检验假设

$$\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q \tag{1}$$

的一些贝叶斯结果, 正如 10.2 节中一样. 在备择假设下, 令

$$\left[\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g\right] = \left[(\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g y^{(g)}, (\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g)^{-1}\right], \quad g = 1, \cdots, q, \tag{2}$$

其中 $p\times r_g$ 阶的矩阵 $\boldsymbol{C}_g$ 有正比于 $|\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}$ 的密度, $n_g = N_g - 1$, 给定 $\boldsymbol{C}_g$ 时, r_g 元随机向量 $\boldsymbol{y}^{(g)}$ 有条件正态分布, 均值为 $\boldsymbol{0}$, 协方差阵为 $(1/N_g)[\boldsymbol{I}_{r_g} - \boldsymbol{C}_g'(\boldsymbol{I}_p + \boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g]^{-1}$, 且 $(\boldsymbol{C}_1, \boldsymbol{y}^{(1)}), \cdots, (\boldsymbol{C}_q, \boldsymbol{y}^{(q)})$ 是独立的. 正如我们将看到的, 我们需要选择适当的整数 $r_1, \cdots, r_q$. 注意到如果 $n_g \geqslant p + r_g$, 则 $|\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}$ 的积分是有限的. 则贝叶斯比的分子是

$$\begin{aligned}
&\text{const} \prod_{g=1}^{q} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g'|^{\frac{1}{2}N_g} \\
&\cdot \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N_g}\left[\boldsymbol{x}_\alpha^{(g)} - (\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\boldsymbol{y}^{(g)}\right]'\right. \\
&\left.\cdot(\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')\left[\boldsymbol{x}_\alpha^{(g)} - (\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\boldsymbol{y}^{(g)}\right]\right\} \\
&\cdot|\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}\left|\boldsymbol{I} - \boldsymbol{C}_g'(\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g\right|^{\frac{1}{2}} \\
&\cdot\exp\left\{-\frac{1}{2}N_g\boldsymbol{y}^{(g)'}\left[\boldsymbol{I} - \boldsymbol{C}_g'(\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')^{-1}\boldsymbol{C}_g^{(g)}\right]\boldsymbol{y}^{(g)}\right\}\mathrm{d}\boldsymbol{y}^{(g)}\mathrm{d}\boldsymbol{C}_g \\
=&\text{const} \prod_{g=1}^{q} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)'}(\boldsymbol{I}+\boldsymbol{C}_g\boldsymbol{C}_g')\boldsymbol{x}_\alpha^{(g)}\right.\right. \\
&\left.\left.-2\boldsymbol{y}^{(g)'}\boldsymbol{C}_g'\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)} + N_g\boldsymbol{y}^{(g)'}\boldsymbol{y}^{(g)}\right]\right\}\mathrm{d}\boldsymbol{y}^{(g)}\mathrm{d}\boldsymbol{C}_g \\
=&\text{const} \prod_{g=1}^{q} \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)'}\boldsymbol{x}_\alpha^{(g)}\right\}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}
\end{aligned} \tag{3}$$

$$\cdot \exp\left\{-\frac{1}{2}N_g\left(\boldsymbol{y}^{(g)}-\boldsymbol{C}_g'\bar{\boldsymbol{x}}^{(g)}\right)'\left(\boldsymbol{y}^{(g)}-\boldsymbol{C}_g'\bar{\boldsymbol{x}}^{(g)}\right)-\frac{1}{2}\operatorname{tr}\boldsymbol{C}_g'\boldsymbol{A}_g\boldsymbol{C}_g\right\}\mathrm{d}\boldsymbol{y}^{(g)}\mathrm{d}\boldsymbol{C}_g$$
$$=\text{const}\prod_{g=1}^{q}\exp\left\{-\frac{1}{2}\left[\operatorname{tr}\boldsymbol{A}_g+N_g\boldsymbol{x}^{(g)'}\boldsymbol{x}^{(g)}\right]\right\}|\boldsymbol{A}_g|^{-\frac{1}{2}r_g}.$$

在原假设下, 令

$$\left[\boldsymbol{\mu}^{(g)},\boldsymbol{\Sigma}_g\right]=\left[(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{y}^{(g)},(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\right],\tag{4}$$

其中 $p\times r$ 矩阵 $\boldsymbol{C}$ 有正比于 $|\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}n}$ 的密度, $n=\sum_{g=1}^{q}n_g$, 给定 $\boldsymbol{C}$ 时, r 元随机向量 $\boldsymbol{y}^{(g)}$ 有条件正态分布, 均值为 $\mathbf{0}$, 协方差阵为 $(1/N_g)[\boldsymbol{I}_r-\boldsymbol{C}'(\boldsymbol{I}_p+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}]^{-1}$, 且 $\boldsymbol{y}^{(1)},\cdots,\boldsymbol{y}^{(g)}$ 是条件独立的. 注意到, 如果 $n\geqslant p+r$, 则 $|\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}n}$ 的积分是有限的. 贝叶斯比的分母是

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{g=1}^{q}\left[|\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}'|^{\frac{1}{2}N_g}\exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N_g}\left[\boldsymbol{x}_\alpha^{(g)}-(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{y}^{(g)}\right]'\right.\right.\tag{5}$$
$$\left.\cdot(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')\left[\boldsymbol{x}_\alpha^{(g)}-(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{y}^{(g)}\right]\right\}$$
$$\cdot|\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}n_g}\left|\boldsymbol{I}-\boldsymbol{C}'(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\right|^{\frac{1}{2}}$$
$$\left.\cdot\exp\left\{-\frac{1}{2}N_g\boldsymbol{y}^{(g)'}\left[\boldsymbol{I}-\boldsymbol{C}'(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}\right]\boldsymbol{y}^{(g)}\right\}d\boldsymbol{y}^{(g)}\right]\mathrm{d}\boldsymbol{C}$$
$$=\text{const}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{g=1}^{q}\exp\left\{-\frac{1}{2}\left[\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)'}(\boldsymbol{I}+\boldsymbol{C}\boldsymbol{C}')\boldsymbol{x}_\alpha^{(g)}\right.\right.$$
$$\left.\left.-2\boldsymbol{y}^{(g)'}\boldsymbol{C}'\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)}+N_g\boldsymbol{y}^{(g)'}\boldsymbol{y}^{(g)}\right]\right\}\mathrm{d}\boldsymbol{y}^{(g)}\mathrm{d}\boldsymbol{C}$$
$$=\text{const}\exp\left\{-\frac{1}{2}\sum_{g=1}^{q}\sum_{\alpha=1}^{N_g}\boldsymbol{x}_\alpha^{(g)'}\boldsymbol{x}_\alpha^{(g)}\right\}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}$$
$$\cdot\exp\left\{-\frac{1}{2}\sum_{g=1}^{q}N_g(\boldsymbol{y}^{(g)}-\boldsymbol{C}'\bar{\boldsymbol{x}}^{(g)})'(\boldsymbol{y}^{(g)}-\boldsymbol{C}'\bar{\boldsymbol{x}}^{(g)})-\frac{1}{2}\operatorname{tr}\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}\right\}\prod_{g=1}^{q}\mathrm{d}\boldsymbol{y}^{(g)}\mathrm{d}\boldsymbol{C}$$
$$=\text{const}\exp\left\{-\frac{1}{2}\left(\operatorname{tr}\boldsymbol{A}+\sum_{g=1}^{q}N_g\bar{\boldsymbol{x}}^{(g)'}\bar{\boldsymbol{x}}^{(g)}\right)\right\}|\boldsymbol{A}|^{-\frac{1}{2}r}.$$

如果

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}r}}{\prod_{g=1}^{q}|\boldsymbol{A}_g|^{\frac{1}{2}r_g}}\geqslant c,\tag{6}$$

则贝叶斯检验方法拒绝原假设. 为了不变性我们希望 $\sum_{g=1}^{q} r_g = r$.

对于 $r_1, \cdots, r_q$ 的选择的约束条件是 $r_g \leqslant n_g - p$, $g = 1, \cdots, q$. 在一些特殊的情况下可以选择 $r_1, \cdots, r_q$ 使得 $(r_1, \cdots, r_q)$ 和 $(N_1, \cdots, N_q)$ 成正比来得到似然比检验, 或者和 $(n_1, \cdots, n_q)$ 成正比来得到修正的似然比检验, 但是由于 $(r_1, \cdots, r_q)$ 必须是正数, 所以有时找不到满足两者中任何一种的解. 下面我们考虑这种方法的推广, 其中包括数 $t_1, \cdots, t_q, t$ 以及 $r_1, \cdots, r_q, r$ 的选择.

假设 $2(p-1) < n_g$, $g = 1, \cdots, q$, 并且取 $r_g \geqslant p$. 令 t_g 是满足 $2p - 1 < r_g + t_g + p < n_g + 1$ 的实数, 且令 t 是满足 $2p - 1 < r + t + p < n + 1$ 的实数. 在备择假设下令 $\boldsymbol{C}_g$ 的边缘密度和 $|\boldsymbol{C}_g\boldsymbol{C}_g'|^{\frac{1}{2}t_g}|\boldsymbol{I} + \boldsymbol{C}_g\boldsymbol{C}_g'|^{-\frac{1}{2}n_g}$ 成正比, $g = 1, \cdots, q$, 在原假设下令 $\boldsymbol{C}$ 的边缘密度和 $|\boldsymbol{C}\boldsymbol{C}'|^{\frac{1}{2}t}|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}n}$ 成正比. (对于 $t_1, \cdots, t_q$ 和 t 的条件是为了确保给定的密度有有限的积分, 参见习题 10.18.) 如果

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(r+t)}}{\prod_{g=1}^{q} |\boldsymbol{A}_g|^{\frac{1}{2}(r_g+t_g)}} \geqslant c, \tag{7}$$

则贝叶斯方法拒绝原假设. 为了不变性, 我们希望 $t = \sum_{g=1}^{q} t_g$. 如果 $t_1, \cdots, t_q$ 满足 $r_g + t_g = kN_g$ 且对于某个 k, $p - 1 < kN_g < N_g - p$, $g = 1, \cdots, q$, 则 (7) 式是似然比检验; 如果 $t_1, \cdots, t_q$ 满足 $r_g + t_g = kn_g$ 且对于某个 k, $p - 1 < kn_g < n_g + 1 - p$, $g = 1, \cdots, q$, 则 (7) 式是修正的似然比检验 [即 $(p-1)/\min_g N_g < k < 1 - p/\min_g N_g$].

定理 10.10.1 如果 $2p < N_g + 1$, $g = 1, \cdots, q$, 则原假设 (1) 的似然比检验和修正的似然比检验是容许的.

现在考虑假设

$$\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(q)}, \quad \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_q. \tag{8}$$

备择假设以前已经处理过. 对于这个原假设, 令

$$[\boldsymbol{\mu}^{(g)}, \boldsymbol{\Sigma}_g] = [(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')\boldsymbol{C}\boldsymbol{y}, (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}], \tag{9}$$

其中 $p \times r$ 矩阵 $\boldsymbol{C}$ 有正比于 $|\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}'|^{-\frac{1}{2}(N-1)}$ 的密度且给定 $\boldsymbol{C}$ 时 r 元向量 $\boldsymbol{y}$ 有条件正态分布, 均值为 $\boldsymbol{0}$, 协方差阵为 $(1/N)[\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{I} + \boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C}]^{-1}$. 如果

$$\frac{|\sum_{g=1}^{q} \boldsymbol{A}_g + \sum_{g=1}^{q} N_g(\bar{\boldsymbol{y}}^{(g)} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}^{(g)} - \bar{\boldsymbol{y}})'|^{\frac{1}{2}r}}{\prod_{g=1}^{q} |\boldsymbol{A}_g|^{\frac{1}{2}r_g}} \geqslant c, \tag{10}$$

则贝叶斯方法拒绝原假设 (8). 如果 $2p < N_g + 1$, $g = 1, \cdots, q$, 先验分布可以像前面那样修正来得到似然比检验和修正的似然比检验.

定理 10.10.2 如果 $2p < N_g + 1$, $g = 1, \cdots, q$, 原假设 (8) 的似然比检验和修正的似然比检验是容许的.

更多的细节可以参考 Kiefer and Schwartz(1965).

10.11 椭球等高分布族

10.11.1 椭球等高观测

令 $\boldsymbol{x}_\alpha^{(g)}$, $\alpha=1,\cdots,N_g$, 是 $\boldsymbol{X}^{(g)}$ 的 N_g 个观测, $\boldsymbol{X}^{(g)}$ 有密度

$$|\boldsymbol{\Lambda}_g|^{-\frac{1}{2}}g\left[(\boldsymbol{x}-\boldsymbol{\nu}^{(g)})'\boldsymbol{\Lambda}_g^{-1}(\boldsymbol{x}-\boldsymbol{\nu}^{(g)})\right], \tag{1}$$

其中 $\mathrm{E}[(\boldsymbol{x}-\boldsymbol{\nu}^{(g)})'\boldsymbol{\Lambda}_g^{-1}(\boldsymbol{x}-\boldsymbol{\nu}^{(g)})]^2=\mathrm{E}(R_g^4)<\infty$, $g=1,\cdots,q$. 需要注意, 同样的函数 $g(\cdot)$ 用来作为 q 个总体的密度. 定义 N, $\boldsymbol{A}_g$, $g=1,\cdots,q$, 和 $\boldsymbol{A}$ 同 10.2 节的 (2) 式. 令 $\boldsymbol{S}_g=(1/n_g)\boldsymbol{A}_g$, 其中 $n_g=N_g-1$, $\boldsymbol{S}=(1/n)\boldsymbol{A}$, 其中 $n=\sum_{g=1}^q n_g$.

由于似然比准则 λ_1 在变换 $\boldsymbol{X}^{(g)}=\boldsymbol{C}\boldsymbol{X}^{(g)}+\boldsymbol{\nu}^{(g)}$ 下是不变的, 在原假设下我们可以取 $\boldsymbol{\Sigma}_1=\cdots=\boldsymbol{\Sigma}_q=\boldsymbol{I}$ 和 $\boldsymbol{\nu}^{(1)}=\cdots=\boldsymbol{\nu}^{(g)}=\boldsymbol{0}$. 则

$$\begin{aligned}
-2\ln\lambda_1&=-\left[\sum_{g=1}^q N_g\ln|\hat{\boldsymbol{\Sigma}}_{g\Omega}|-N\ln|\hat{\boldsymbol{\Sigma}}_\omega|\right] \qquad (2)\\
&=-\left[\sum_{g=1}^g N_g\ln\left|\boldsymbol{I}+\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)\right|-N\ln\left|\boldsymbol{I}+\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)\right|\right]\\
&=-\left\{\sum_{g=1}^q N_g\left[\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)-\frac{1}{2}\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)^2+O_p\left(N_g^{-3}\right)\right]\right.\\
&\qquad\left.-N\left[\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)-\frac{1}{2}\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)^2+O_p\left(N^{-3}\right)\right]\right\}\\
&=\frac{1}{2}\sum_{g=1}^q N_g\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)^2-\frac{1}{2}N\mathrm{tr}\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)^2+O_p\left(N^{-3}\right)\\
&=\frac{1}{2}\sum_{g=1}^q N_g\left[\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)\right]'\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I}\right)\\
&\qquad-\frac{1}{2}N\left[\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)\right]'\mathrm{vec}\left(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I}\right)+O_p\left(N^{-3}\right).
\end{aligned}$$

由定理 3.6.2

$$\sqrt{N_g}\mathrm{vec}\,(\boldsymbol{S}_g-\boldsymbol{I}_p)\xrightarrow{d}N\left[\boldsymbol{0},(\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})+\kappa\,\mathrm{vec}\,\boldsymbol{I}_p(\mathrm{vec}\,\boldsymbol{I}_p)'\right], \tag{3}$$

并且 $n_g\boldsymbol{S}_g=N_g\hat{\boldsymbol{\Sigma}}_{g\Omega}\,(g=1,\cdots,q)$ 是独立的. 令 $N_g=k_gN$, $g=1,\cdots,q$, $\sum_{g=1}^q k_g=1$, 且令 $N\to\infty$. 在渐近理论中 $\mathrm{vec}\,(\boldsymbol{S}_1-\boldsymbol{I}),\cdots,\mathrm{vec}\,(\boldsymbol{S}_q-\boldsymbol{I})$ 的极限分布和 8.8 节中 $\bar{\boldsymbol{y}}^{(1)},\cdots,\bar{\boldsymbol{y}}^{(q)}$ 的分布相同, 只是把 8.8 节中的 $\boldsymbol{\Sigma}$ 用 $(\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})+\kappa\,\mathrm{vec}\,\boldsymbol{I}_p(\mathrm{vec}\,\boldsymbol{I}_p)'$ 来替代.

当 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 时, $\sqrt{N_g}(s_{ii}^{(g)}-1)$ 的极限分布的方差是 $3\kappa+2$; $\sqrt{N_g}(s_{ii}^{(g)}-1)$ 和 $\sqrt{N_g}(s_{jj}^{(g)}-1)\,(i\neq j)$ 的极限分布的协方差是 κ; $s_{ij}^{(g)}\,(i\neq j)$ 的方差是 $\kappa+1$; 集合 $\sqrt{N}(s_{11}^{(g)}-1),\cdots,\sqrt{N}(s_{pp}^{(g)}-1)$ 和集合 $(s_{ij}^{(g)})\,(i\neq j)$ 是独立的; 并且 $s_{ij}^{(g)}\,(i<j)$ 是相互不相关的 (正如 7.9.1 节中一样).

令 $\bar{\boldsymbol{y}}_g=\operatorname{vec}(\hat{\boldsymbol{\Sigma}}_{g\Omega}-\boldsymbol{I})$, $\bar{\boldsymbol{y}}=\operatorname{vec}(\hat{\boldsymbol{\Sigma}}_\omega-\boldsymbol{I})$. 则 $\bar{\boldsymbol{y}}=\sum_{g=1}^q(N_g/N)\bar{\boldsymbol{y}}_g$ 且

$$\begin{aligned}-2\ln\lambda&=\frac{1}{2}\sum_{g=1}^q N_g(\bar{\boldsymbol{y}}_g-\bar{\boldsymbol{y}})'(\bar{\boldsymbol{y}}_g-\bar{\boldsymbol{y}})\\&=\operatorname{tr}\frac{1}{2}\sum_{g=1}^q N_g(\bar{\boldsymbol{y}}_g-\bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_g-\bar{\boldsymbol{y}})'\\&=\frac{1}{2}\operatorname{tr}\left(\sum_{g=1}^q N_g\bar{\boldsymbol{y}}_g\bar{\boldsymbol{y}}_g'-N\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}'\right).\end{aligned}\tag{4}$$

令 $\boldsymbol{Q}$ 是一个 $q\times q$ 正交矩阵, 且最后一列是 $(\sqrt{N_1/N},\cdots,\sqrt{N_q/N})'$. 定义

$$(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_q)=\left(\sqrt{N_1}\bar{\boldsymbol{y}}_1,\cdots,\sqrt{N_g}\bar{\boldsymbol{y}}_g\right)\boldsymbol{Q}.\tag{5}$$

则 $\boldsymbol{w}_q=\sqrt{N}\bar{\boldsymbol{y}}$ 且

$$\sum_{g=1}^q N_g\bar{\boldsymbol{y}}_g\bar{\boldsymbol{y}}_g'-N\bar{\boldsymbol{y}}\bar{\boldsymbol{y}}=\sum_{g=1}^{q-1}\boldsymbol{w}_g\boldsymbol{w}_g'.\tag{6}$$

在这些项中,

$$-2\ln\lambda_1=\frac{1}{2}\sum_{g=1}^{q-1}\boldsymbol{w}_g'\boldsymbol{w}_g+O_p(N^{-3}),\tag{7}$$

$\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{q-1}$ 是渐近独立的, $\boldsymbol{w}_g$ 和 $\sqrt{N}\bar{\boldsymbol{y}}_g$ 有相同的协方差阵, 即 $(\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})+\kappa\operatorname{vec}\boldsymbol{I}_p(\operatorname{vec}\boldsymbol{I}_p)$. 则 $\boldsymbol{w}_g'\boldsymbol{w}_g=\sum_{i,j=1}^p(w_{ij}^{(g)})^2=\sum_{i=1}^p(w_{ii}^{(g)})^2+2\sum_{i<j}(w_{ij}^{(g)})^2$. $w_{11}^{(g)},\cdots,w_{pp}^{(g)}$ 的协方差阵是 $2(\kappa+1)\boldsymbol{I}_p+\kappa\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}$, 其中 $\boldsymbol{\varepsilon}=(1,\cdots,1)'$. 这个矩阵的特征根包括重数为 $p-1$ 的根 $2(\kappa+1)$ 和单根 $2(\kappa+1)+p\kappa$. 因此 $\sum_{i=1}^p(w_{ii}^{(g)})^2$ 有分布 $2(\kappa+1)\chi_{p-1}^2+[2(\kappa+1)+p\kappa]\chi_1^2$. $2\sum_{i<j}(w_{ij}^{(g)})^2$ 的分布是 $2(\kappa+1)\chi_{p(p-1)/2}^2$.

定理 10.11.1　*当从 (1) 中抽样且原假设为真时,*

$$-2\ln\lambda_1\xrightarrow{d}(\kappa+1)\chi_{(q-1)(p-1)(p+2)/2}^2+[(\kappa+1)+p\kappa/2]\chi_{q-1}^2.\tag{8}$$

当 $\kappa=0$ 时, $-2\ln\lambda_1\xrightarrow{d}\chi_{(q-1)p(p+1)/2}$ 和 10.5 节中的 (12) 式一致. 从 10.4 节中得出的分布的有效性依赖于观测服从正态分布, 定理 10.11.1 表明渐近理论甚至依赖于非正态性.

检验 10.3 节中原假设 (2) 的似然比准则是 $\lambda_1\lambda_2$ 或 V_1V_2. 引理 10.4.1 表示在正态条件下 V_1 和 V_2(或等价地 λ_1 和 λ_2) 是独立的. 在椭球等高的情况下我们想证明 $\ln V_1$ 和 $\ln V_2$ 是渐近独立的.

引理 10.11.1 令 $\boldsymbol{A}_1 = n_1\boldsymbol{S}_1$ 和 $\boldsymbol{A}_2 = n_2\boldsymbol{S}_2$ 由 10.2 节中的 (2) 式来定义, 其中 $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{I}$. 则 $\boldsymbol{A}_1(\boldsymbol{A}_1+\boldsymbol{A}_2)^{-1}$ 和 $\boldsymbol{A}_1+\boldsymbol{A}_2$ 是渐近独立的

证明 令 $(1/\sqrt{n_g})(\boldsymbol{A}_g - n_g\boldsymbol{I}) = \boldsymbol{W}_g$, $g=1,2$. 则

$$\sqrt{n_1}\left[\boldsymbol{A}_1(\boldsymbol{A}_1+\boldsymbol{A}_2)^{-1} - \frac{n_1}{n_1+n_2}\boldsymbol{I}\right] = \frac{n_1n_2}{(n_1+n_2)^2}\boldsymbol{w}_1 - \frac{n_1\sqrt{n_1n_2}}{(n_1+n_2)^2}\boldsymbol{w}_2 + O_p(1), \quad (9)$$

$$\sqrt{n_1+n_2}[\boldsymbol{A}_1+\boldsymbol{A}_2-(n_1+n_2)\boldsymbol{I}] = \sqrt{\frac{n_1}{n_1+n_2}}\boldsymbol{w}_1 + \sqrt{\frac{n_2}{n_1+n_2}}\boldsymbol{w}_2 + O_p(1). \quad (10)$$

则

$$\mathrm{E}\left\{\mathrm{vec}\left(\frac{n_1n_2}{(n_1+n_2)^2}\boldsymbol{w}_1 - \frac{n_1\sqrt{n_1n_2}}{(n_1+n_2)^2}\boldsymbol{w}_2\right)\right. \quad (11)$$

$$\left.\left[\mathrm{vec}\left(\sqrt{\frac{n_1}{n_1+n_2}}\boldsymbol{w}_1 + \sqrt{\frac{n_2}{n_1+n_2}}\boldsymbol{w}_2\right)\right]'\right\} = \boldsymbol{0}. \qquad \blacksquare$$

依次利用引理 10.11.1 于 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_1+\boldsymbol{A}_2$, $\boldsymbol{A}_1+\boldsymbol{A}_2$ 和 $\boldsymbol{A}_1+\boldsymbol{A}_2+\boldsymbol{A}_3$, 等等, 我们得到 $\boldsymbol{A}_1\boldsymbol{A}^{-1}, \boldsymbol{A}_2\boldsymbol{A}^{-1}, \cdots, \boldsymbol{A}_q\boldsymbol{A}^{-1}$ 和 $\boldsymbol{A} = \boldsymbol{A}_1 + \cdots + \boldsymbol{A}_q$ 是独立的. 从而 V_1 和 V_2 是渐近独立的.

定理 10.11.2 当 $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_g$ 且 $\boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(g)}$ 时,

$$\begin{aligned} -2\ln\lambda_1\lambda_2 &= -2\ln\lambda_1 - 2\ln\lambda_2 \\ &\xrightarrow{d} (\kappa+1)\chi^2_{(q-1)(p-1)(p+2)/2} + [(\kappa+1)+p\kappa/2]\chi^2_{q-1} + \chi^2_{p(q-1)}. \end{aligned} \quad (12)$$

球形假设是 $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$(或者 $\boldsymbol{\Lambda} = \lambda\boldsymbol{I}$). 准则是 $\lambda_1\lambda_2$, 其中

$$\lambda_1 = \left(\frac{|\boldsymbol{A}|}{\prod_{i=1}^p a_{ii}}\right)^{N/2}, \quad \lambda_2 = \left[\frac{\prod_{i=1}^p a_{ii}}{\left(\frac{\mathrm{tr}\,\boldsymbol{A}}{p}\right)^p}\right]^{N/2}. \quad (13)$$

第一个因子是 $\boldsymbol{X}$ 的分量独立的准则, 而第二个是分量的方差相等的准则. 对于第一个我们在定理 9.10 中固定 $q=p$ 和 $p_i=1$, 对于第二个我们固定 $q=p$ 和 $p=1$. 因此

$$-2\ln(\lambda_1\lambda_2) \xrightarrow{d} (1+\kappa)\chi^2_{p-1} + \frac{1}{2}(3\kappa+2)\chi^2_{p-1}. \quad (14)$$

10.11.2 椭球等高矩阵分布

考虑密度

$$\begin{aligned} &\prod_{g=1}^q |\boldsymbol{\Lambda}_g|^{-N_g/2} g\left[\mathrm{tr}\sum_{g=1}^q \boldsymbol{\Lambda}_g^{-1}\left(\boldsymbol{X}^{(g)} - \boldsymbol{\nu}_g\boldsymbol{\varepsilon}'_{N_g}\right)\left(\boldsymbol{X}^{(g)} - \boldsymbol{\nu}_g\boldsymbol{\varepsilon}'_{N_g}\right)'\right] \\ =&\prod_{g=1}^q |\boldsymbol{\Lambda}_g|^{-N_g/2} g\left[\mathrm{tr}\sum_{g=1}^q \boldsymbol{\Lambda}_g^{-1}\boldsymbol{\Lambda}_g + \sum_{g=1}^q N_g(\bar{\boldsymbol{x}}^{(g)} - \boldsymbol{\nu}^{(g)})'\boldsymbol{\Lambda}_g^{-1}(\bar{\boldsymbol{x}}^{(g)} - \boldsymbol{\nu}^{(g)})\right]. \end{aligned} \quad (15)$$

在这个密度中, $(\boldsymbol{A}_g, \bar{\boldsymbol{x}}_g)$, $g=1,\cdots,q$ 是统计量的一个充分集, 且似然比准则是 10.2 节中的 (8) 式, 和正态性的相同 [Anderson and Fang(1990b)].

定理 10.11.3 令 $\boldsymbol{f}(\boldsymbol{X})$ 是 $\boldsymbol{X}=(\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)})\,(p\times N)$ 的一个向量值函数, 满足对于每一个 $(\boldsymbol{\nu}^{(1)},\cdots,\boldsymbol{\nu}^{(q)})$

$$\boldsymbol{f}(\boldsymbol{X}^{(1)}+\boldsymbol{\nu}^{(1)}\boldsymbol{\varepsilon}'_{N_1},\cdots,\boldsymbol{X}^{(q)}+\boldsymbol{\nu}^{(q)}\boldsymbol{\varepsilon}'_{N_q})=\boldsymbol{f}(\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)}) \tag{16}$$

且对于每一个非奇异矩阵 $\boldsymbol{C}$

$$\boldsymbol{f}(\boldsymbol{C}\boldsymbol{X}^{(1)},\cdots,\boldsymbol{C}\boldsymbol{X}^{(q)})=\boldsymbol{f}(\boldsymbol{X}^{(1)},\cdots,\boldsymbol{X}^{(q)}). \tag{17}$$

则对于有任意密度 (15) 且 $\boldsymbol{\Lambda}_1=\cdots=\boldsymbol{\Lambda}_q$ 的 $\boldsymbol{X}$, $\boldsymbol{f}(\boldsymbol{X})$ 的分布和当 $\boldsymbol{X}$ 有正态密度 (15) 时的 $\boldsymbol{f}(\boldsymbol{X})$ 的分布相同.

定理 10.11.3 的证明和定理 4.5.4 的证明类似. 这个定理说明, 当 $\boldsymbol{X}$ 有密度 (15) 且 $\boldsymbol{\Lambda}_1=\cdots=\boldsymbol{\Lambda}_q$ 时 10.2 节中 (10) 式的准则 V_1 的分布和 $\boldsymbol{X}$ 为正态时相同. 因此分布和它们的渐近展开就是 10.4 节和 10.5 节中讨论的那样.

推论 10.11.1 令 $\boldsymbol{f}(\boldsymbol{X})$ 是 $\boldsymbol{X}\,(p\times N)$ 的向量值函数, 满足对于每一个 $\boldsymbol{\nu}$

$$\boldsymbol{f}(\boldsymbol{X}+\boldsymbol{\nu}\boldsymbol{\varepsilon}'_N)=\boldsymbol{f}(\boldsymbol{X}) \tag{18}$$

且 (17) 式成立. 则当 $\boldsymbol{X}$ 有任意密度 (15) 且 $\boldsymbol{\Lambda}_1=\cdots=\boldsymbol{\Lambda}_q$ 和 $\boldsymbol{\nu}^{(1)}=\cdots=\boldsymbol{\nu}^{(q)}$ 时, $\boldsymbol{f}(\boldsymbol{X})$ 的分布和当 $\boldsymbol{X}$ 有正态密度 (15) 时 $\boldsymbol{f}(\boldsymbol{X})$ 的分布相同.

从而 10.3 节中 (7) 式的准则 λ 或 (11) 式的准则 V 的分布与密度 (15) 相同, 正如 $\boldsymbol{X}$ 服从正态分布时一样.

令 $\boldsymbol{X}\,(p\times N)$ 有密度

$$|\boldsymbol{\Lambda}|^{-N/2}g\left[\operatorname{tr}\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{\nu}\boldsymbol{\varepsilon}'_N)(\boldsymbol{X}-\boldsymbol{\nu}\boldsymbol{\varepsilon}'_N)'\right]. \tag{19}$$

则对于某个 $\lambda>0$, 检验原假设 $\boldsymbol{\Lambda}=\lambda\boldsymbol{I}$ 的似然比准则是 10.7 节的 (7) 式, 且在原假设下它的分布和 $\boldsymbol{X}$ 服从正态分布时的相同.

更多的细节请参考 Anderson and Fang(1990b) 及 Fang and Zhang(1990).

习 题

10.1 (10.2 节) 四个测量的协方差阵如下 (从表 3.4 得到). 总体是变色鸢尾 (1), 山鸢尾 (2), 维吉尼亚鸢尾 (3), 每个样本包含 50 个观测.

$$\boldsymbol{A}_1=\begin{pmatrix}13.0552 & 4.1740 & 8.9620 & 2.7332\\ 4.1740 & 4.8250 & 4.0500 & 2.0190\\ 8.9620 & 4.0500 & 10.8200 & 3.5820\\ 2.7332 & 2.0190 & 3.5820 & 1.9162\end{pmatrix},$$

$$A_2 = \begin{pmatrix} 6.0882 & 4.8616 & 0.8014 & 0.5062 \\ 4.8616 & 7.0408 & 0.5732 & 0.4556 \\ 0.8014 & 0.5732 & 1.4778 & 0.2974 \\ 0.5062 & 0.4556 & 0.2974 & 0.5442 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 19.8128 & 4.5944 & 14.8612 & 2.4056 \\ 4.5944 & 5.0962 & 3.4976 & 2.3338 \\ 14.8612 & 3.4976 & 14.9248 & 2.3924 \\ 2.4056 & 2.3338 & 2.3924 & 3.6962 \end{pmatrix}.$$

(a) 在显著性水平 5% 下检验假设 $\Sigma_1 = \Sigma_2$.

(b) 在显著性水平 5% 下检验假设 $\Sigma_1 = \Sigma_2 = \Sigma_3$.

10.2 (10.2 节)

(a) 令 $Y^{(g)}\,(g = 1, \cdots, q)$ 是一个 p 元随机向量集. 假设

$$\mathrm{E}(Y^{(g)}) = 0, \quad \mathrm{E}(Y^{(g)}Y^{(h)\prime}) = \delta_{gh}\Sigma_g.$$

令 C 是一个 q 阶正交矩阵且其最后一行的每个元素都是

$$c_{qh} = 1/\sqrt{q}.$$

定义

$$\mathbf{Z}^{(g)} = \sum_{h=1}^{q} c_{gh} Y^{(h)}, \quad g = 1, \cdots, q.$$

证明

$$\mathrm{E}(\mathbf{Z}^{(q)}\mathbf{Z}^{(g)\prime}) = \mathbf{0}, \quad g = 1, \cdots, q-1,$$

当且仅当

$$\Sigma_1 = \Sigma_2 = \cdots = \Sigma_q.$$

(b) 令 $X_\alpha^{(g)}\,(\alpha = 1, \cdots, N)$ 是从正态总体 $N(\mu^{(g)}, \Sigma_g)\,(g = 1, \cdots, q)$ 中得到的随机样本. 基于 $\mathbf{Z}^{(q)}$ 和集合 $\mathbf{Z}^{(1)}, \cdots, \mathbf{Z}^{(q-1)}$ 的独立性检验并且利用 (a) 中的结果来构造假设

$$H: \Sigma_1 = \cdots = \Sigma_q$$

的检验. 找出在 $p = 2$ 的情况下准则的精确分布.

10.3 (10.2 节) 关于 $\sigma_1^2 = \sigma_2^2$ 的修正的似然比检验的无偏性. 证明 (14) 式是无偏的. [提示: 令 $G = n_1F/n_2$, $r = \sigma_1^2/\sigma_2^2$, 且 $c_1 < c_2$ 是 $G^{\frac{1}{2}n_2}(1+G)^{-\frac{1}{2}(n_1+n_2)} = k$ 的两个解, 即修正的似然比准则的判别值. 则

$$\begin{aligned}\Pr\{\text{接受} \mid \sigma_1^2/\sigma_2^2 = r\} &= \text{const}\int_{c_1}^{c_2} r^{\frac{1}{2}n_1} G^{\frac{1}{2}n_1 - 1}(1 + rG)^{-\frac{1}{2}(n_1+n_2)}\mathrm{d}G \\ &= \text{const}\int_{rc_1}^{rc_2} H^{\frac{1}{2}n_1 - 1}(1 + H)^{-\frac{1}{2}(n_1+n_2)}\mathrm{d}H.\end{aligned}$$

证明上式关于 r 的导数当 $0 < r < 1$ 时正的, 当 $r = 1$ 时是 0, 当 $r > 1$ 时是负的.]

10.4 (10.2 节) 证明 (19) 式的极限分布是 χ_f^2, 其中 $f = \frac{1}{2}p(p+1)(q-1)$. [提示: 令 $\Sigma = I$. 证明 (19) 式的极限分布是

$$\frac{1}{2}\sum_{i=1}^{p}\sum_{g=1}^{q}n_g(s_{ii}^{(g)}-s_{ii})^2+\sum_{i<j}\sum_{g=1}^{q}n_g(s_{ij}^{(g)}-s_{ij})^2$$

的极限分布, 其中 $\boldsymbol{S}^{(g)}=(s_{ij}^{(g)})$, $\boldsymbol{S}=(s_{ij})$, 且 $\sqrt{n_g}(s_{ij}^{(g)}-\delta_{ij})\ (i\leqslant j)$ 在极限分布中是独立的, $\sqrt{n_g}(s_{ii}^{(g)}-1)$ 的极限分布是 $N(0,2)$, $\sqrt{n_g}s_{ij}^{(g)}\ (i<j)$ 的极限分布是 $N(0,1)$.]

10.5 (10.4 节) 利用 Wishart 密度的积分证明 (15) 式.

[提示: $\mathrm{E}(V_1^h)=\mathrm{E}(\prod_{g=1}^{q}|\boldsymbol{A}_g|^{\frac{1}{2}n_g}|\boldsymbol{A}|^{-\frac{1}{2}n})$ 可以写成一个常数与 $|\boldsymbol{A}|^{-\frac{1}{2}n}\prod_{g=1}^{q}w(\boldsymbol{A}_g|\boldsymbol{\Sigma},n_g+hn_g)$ 之乘积的积分. 关于 $\sum_{g=1}^{q}\boldsymbol{A}_g=\boldsymbol{A}$ 积分得到一个常数乘以 $w(\boldsymbol{A}|\boldsymbol{\Sigma},n)$.]

10.6 (10.4 节) 利用 Wishart 密度和正态密度的积分证明 (16) 式. [提示: $\sum_{g=1}^{q}N_g(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}^{(g)}-\bar{\boldsymbol{x}})'$ 和 $\sum_{f=1}^{q-1}\boldsymbol{y}_f\boldsymbol{y}_f'$ 同分布. 利用习题 10.5 的提示.]

10.7 (10.6 节) 令 $\boldsymbol{x}_1^{(\nu)},\cdots,\boldsymbol{x}_N^{(\nu)}$ 是从正态总体 $N(\boldsymbol{\mu}^{(\nu)},\boldsymbol{\Sigma}\boldsymbol{\nu})\ (\boldsymbol{\nu}=1,2)$ 中得到的观测, 且令 $\boldsymbol{A}_\nu=\sum(\boldsymbol{x}^{(\nu)}-\bar{\boldsymbol{x}}^{(\nu)})(\boldsymbol{x}^{(\nu)}-\bar{\boldsymbol{x}}^{(\nu)})'$.

(a) 证明关于 $H:\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$ 的似然比检验等价于如果

$$T=\frac{|\boldsymbol{A}_1|\cdot|\boldsymbol{A}_2|}{|\boldsymbol{A}_1+\boldsymbol{A}_2|^2}\leqslant C$$

则拒绝 H.

(b) 令 $d_1^2,d_2^2,\cdots,d_p^2$ 是 $|\boldsymbol{\Sigma}_1-\lambda\boldsymbol{\Sigma}_2|=0$ 的根, 且令

$$\boldsymbol{D}=\begin{pmatrix}d_1&0&\cdots&0\\0&d_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&d_p\end{pmatrix}.$$

证明 T 和 $|\boldsymbol{B}_1|\cdot|\boldsymbol{B}_2|\,/\,|\boldsymbol{B}_1+\boldsymbol{B}_2|^2$ 同分布, 其中 $\boldsymbol{B}_1$ 服从 $W(\boldsymbol{D}^2,N-1)$ 且 $\boldsymbol{B}_2$ 服从 $W(\boldsymbol{I},N-1)$. 证明 T 和 $|\boldsymbol{DC}_1\boldsymbol{D}|\cdot|\boldsymbol{C}_2|\,/\,|\boldsymbol{DC}_1\boldsymbol{D}+\boldsymbol{C}_2|^2$ 同分布, 其中 $\boldsymbol{C}_i$ 服从 $W(\boldsymbol{I},N-1)$.

10.8 (10.6 节) 对于 $p=2$ 证明

$$\begin{aligned}\Pr\{V_1\leqslant\nu\}=&I_a(n_1-1,n_2-1)\\&+B^{-1}(n_1-1,n_2-1)\nu^{(n_1+n_2-2)/n}\int_g^b x^{-2n_2/n}(1-x_1)^{-n_1/n}\mathrm{d}x_1\\&+1-I_b(n_1-1,n_2-1),\end{aligned}$$

其中 $a<b$ 是 $x_1^{n_1}(1-x_1)^{n_2}=\nu\leqslant n_1^{n_1}n_2^{n_2}/n^n$ 的两个根. [提示: 对由 (8) 式定义的密度进行积分即得这个]

10.9 (10.6 节) 对于 $p=2$ 和 $n_1=n_2=m$, 证明

$$\begin{aligned}&\Pr\{V_1\leqslant\nu\}\\&=2I_a(m-1,m-1)+2B^{-1}(m-1,m-1)\nu^{1-(1/m)}\ln\frac{1+\sqrt{1-4\nu^{1/m}}}{1-\sqrt{1-4\nu^{1/m}}},\end{aligned}$$

其中 $a=\frac{1}{2}[1-\sqrt{1-4\nu^{1/m}}]$.

10.10 (10.7 节) 在原假设下求出 $p=2$ 时 W 的分布

(a) 直接用 $\boldsymbol{A}$ 的分布去求,

(b) 从特征根的分布去求 (第 13 章).

10.11 (10.7 节) 令 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是从总体 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 中得到的样本. 问检验假设 $\boldsymbol{\mu} = k\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0$ 的似然比准则是什么, 其中 $\boldsymbol{\mu}_0$ 和 $\boldsymbol{\Sigma}_0$ 是确定的而 k 是不定的.

10.12 (10.7 节) 令 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{N_1}^{(1)}$ 是从总体 $N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_1)$ 中得到的样本, $\boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_{N_2}^{(2)}$ 是从总体 $N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_2)$ 中得到的样本. 问检验假设 $\boldsymbol{\Sigma}_1 = k^2\boldsymbol{\Sigma}_2$ 的似然比准则是什么, 其中 k 是不定的. 检验假设 $\boldsymbol{\mu}^{(1)} = k\boldsymbol{\mu}^{(2)}$ 且 $\boldsymbol{\Sigma}_1 = k^2\boldsymbol{\Sigma}_2$ 的似然比准则是什么, 其中 k 是不定的.

10.13 (10.7 节) 令 $\boldsymbol{x}_\alpha\,(\alpha = 1, \cdots, N)$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 p 元观测. 我们定义如下假设:

$$H : \boldsymbol{\mu} = \boldsymbol{0},\ \boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0,$$

$$H_1 : \boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0,$$

$$H_2 : \boldsymbol{\mu} = \boldsymbol{0}, \quad \text{给定}\boldsymbol{\Sigma} = k^2\boldsymbol{\Sigma}_0.$$

所有情况下 k^2 都是不确定的, 但是 $\boldsymbol{\Sigma}_0$ 是确定的. 找出检验 H_2 的似然比准则 λ_2. 给出在 H_2 下 $-2\ln\lambda_2$ 的渐近分布. 求出在 H_2 下 λ_2 的一个合适的单调函数的精确分布.

10.14 (10.7 节) 找出习题 10.13 中检验 H 的似然比准则 λ (给定 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$). 在 H 下 $-2\ln\lambda$ 的渐近分布是什么?

10.15 (10.7 节) 证明 $\lambda = \lambda_1\lambda_2$, 其中 λ 是习题 10.14 中定义的, λ_2 是习题 10.13 中定义的, 而 λ_1 是习题 10.13 中检验 H_1 的似然比准则. 在 H 下 λ_1 和 λ_2 是独立的吗? 并证明你的结论.

10.16 (10.7 节) 验证 $\operatorname{tr}\boldsymbol{B}\boldsymbol{\Psi}_0^{-1}$ 服从自由度为 $p(N-1)$ 的 χ^2 分布.

10.17 (10.7.1 节) *球形检验的容许性*. 证明球形的似然比检验是容许的. [提示: 在原假设下令 $\boldsymbol{\Sigma} = [1/(1+\eta^2)]\boldsymbol{I}$, 且令 η 有密度 $(1+\eta^2)^{-\frac{1}{2}np}(\eta^2)^{p-\frac{1}{2}}$.]

10.18 (10.10.1 节) 证明当 $r \geqslant p$ 时, 如果 $2p - 1 < t + r + p < n + 1$, 则

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left|\sum_{i=1}^{r}\boldsymbol{x}_i\boldsymbol{x}_i'\right|^{\frac{1}{2}t}\left|\boldsymbol{I} + \sum_{i=1}^{r}\boldsymbol{x}_i\boldsymbol{x}_i'\right|^{-\frac{1}{2}n}\prod_{i=1}^{r}\mathrm{d}\boldsymbol{x}_i < \infty.$$

[提示: 如果 $\boldsymbol{A}$ 是半正定的 $|\boldsymbol{A}|/|\boldsymbol{I}+\boldsymbol{A}| \leqslant 1$. 还有, 如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r$ 独立同分布且服从 $N(\boldsymbol{0}, \boldsymbol{I})$, 则 $|\sum_{i=1}^{r}\boldsymbol{x}_i\boldsymbol{x}_i'|$ 服从分布 $\chi_r^2\chi_{r-1}^2\cdots\chi_{r-p+1}^2$.]

10.19 (10.10.1 节) 证明

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|\boldsymbol{C}\boldsymbol{C}'|^{\frac{1}{2}t}\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}}\mathrm{d}\boldsymbol{C} = \text{const}\,|\boldsymbol{A}|^{-\frac{1}{2}(t+r)},$$

其中 $\boldsymbol{C}$ 是 $p \times r$ 的. [提示: 如果 $\boldsymbol{C}$ 的密度和 $\mathrm{e}^{-\frac{1}{2}\operatorname{tr}\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}}$ 成比, 则 $\boldsymbol{C}\boldsymbol{C}'$ 服从分布 $W(\boldsymbol{A}^{-1}, r)$.]

10.20 (10.10.1 节) 利用习题 10.18 的结果完成定理 10.10.1 的证明.

第11章 主　成　分

11.1 引　言

主成分是方差具有特定性质的随机变量或统计变量的线性组合. 例如, 第一主成分是具有极大方差的正规线性组合 (系数的平方和为 1). 其实, 把原始向量变量转变为主成分向量等同于把坐标轴转到一个具有固有统计性质的新的坐标体系下. 这种坐标体系的选择和以前处理过的许多与坐标系无关的问题明显不同.

主成分其实是协方差阵的特征向量. 因此对主成分的研究可以看作是在统计问题中加入 (对于半正定矩阵的) 特征根和特征向量的一般发展.

从统计理论的角度来看, 主成分的集合生成了一个合适的坐标集, 并且这些成分的方差标识了它们的统计性质. 在实际统计问题中, 主成分的方法是用来找到具有极大方差的线性组合的. 在许多探索性研究中, 由于所考虑的变量的个数太大而很难处理. 由于在这些研究中人们感兴趣的是偏差, 所以一个减少需要处理的变量个数的方法就是舍去具有小方差的线性组合, 而只研究大方差的组合. 例如, 一个人体人类学家对于每一个个体可能会有很多关于长度和宽度的测量, 例如耳朵的长度, 耳朵的宽度, 面部的长度, 面部的宽度等. 人们感兴趣的可能是描述和分析各个个体之间在这些人类特征之间的差别. 最后人们想解释这些差别, 但是首先他需要知道哪些测量或测量的组合可以表示大部分的变差, 也就是哪些应该进一步研究. 主成分给出了一个的新的测量的线性组合集. 可能个体之间的大部分变差留在三个线性组合中; 则人类学家可以直接对这三个量进行研究; 其他的线性组合由于个体之间的差异太小以至于研究它们几乎不能分辨个体变差.

Hotelling(1933) 提出了上述思想的大部分, 并给出了一个相当完整的讨论.

在 11.2 节中我们在具有上面讨论的性质的总体中定义了主成分的概念, 它们定义了一个到对角协方差阵的正交变换. 极大似然估计在样本中具有类似的性质 (11.3 节). 关于计算的简要讨论在 11.4 节中给出, 一个数值例子将在 11.5 节中给出. 样本主成分的系数的渐近分布和样本方差的渐近分布也已给出并且用来得到每个参数的大样本检验和置信区间 (11.6 节); 对于一个协方差阵的特征根的精确置信界也已经得到. 在 11.7 节中我们考虑关于这些根的其他假设检验.

11.2 总体中主成分的定义

假设 p 元随机向量 $\boldsymbol{X}$ 有协方差阵 $\boldsymbol{\Sigma}$. 由于在本章中我们只关注方差和协方差, 所以我们将假设均值向量为 $\boldsymbol{0}$. 并且除了协方差阵外, $\boldsymbol{X}$ 的真实分布与这里用到的思想和代数是不相关的; 但是, 如果 $\boldsymbol{X}$ 是服从正态分布的, 则主成分将会有更多的意义.

在下面的处理中我们不会利用通常的特征根和特征向量的理论. 事实上, 这些理论将会间接得到. 我们也将处理 $\boldsymbol{\Sigma}$ 奇异 (即半正定) 以及 $\boldsymbol{\Sigma}$ 有重根的情形.

令 $\boldsymbol{\beta}$ 是一个 p 元列向量且满足 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$. $\boldsymbol{\beta}'\boldsymbol{X}$ 的方差是

$$\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X})^2=\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta})=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}. \tag{1}$$

为了确定具有极大方差的正规线性组合 $\boldsymbol{\beta}'\boldsymbol{X}$, 我们需要找出满足 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ 且极大化 (1) 的向量 $\boldsymbol{\beta}$. 令

$$\phi=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}-\lambda(\boldsymbol{\beta}'\boldsymbol{\beta}-1)=\sum_{i,j}\beta_i\sigma_{ij}\beta_j-\lambda\left(\sum_i\beta_i^2-1\right), \tag{2}$$

其中 λ 是一个拉格朗日乘数. (利用附录的定理 A.4.3) 偏导数 $(\partial\phi/\partial\beta_i)$ 组成的向量是

$$\frac{\partial\phi}{\partial\boldsymbol{\beta}}=2\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}. \tag{3}$$

由于 $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}'\boldsymbol{\beta}$ 在包含 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ 的区域的任何地方都有导数, 所以极大化 $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ 的向量 $\boldsymbol{\beta}$ 一定使得表达式 (3) 等于 0, 即

$$(\boldsymbol{\Sigma}-\lambda\boldsymbol{I})\boldsymbol{\beta}=0. \tag{4}$$

为了得到 (4) 式的满足 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ 的解, $\boldsymbol{\Sigma}-\lambda\boldsymbol{I}$ 必须是奇异的; 换句话说, λ 必须满足

$$|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|=0. \tag{5}$$

函数 $|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|$ 是 λ 的 p 次多项式. 因此 (5) 式有 p 个根, 令这些根为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$. [在 (6) 式中 $\boldsymbol{\beta}'$ 复共轭证明 λ 是实数.] 如果在 (4) 式两边同时左乘 $\boldsymbol{\beta}'$, 就得到

$$\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}=\lambda\boldsymbol{\beta}'\boldsymbol{\beta}=\lambda. \tag{6}$$

这证明如果 $\boldsymbol{\beta}$ 满足 (4) 式 (且 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$), 则 $\boldsymbol{\beta}'\boldsymbol{X}$ 的方差 [由 (1) 式给出] 是 λ. 因此我们应该用 (4) 式的最大根 λ_1 来表示极大方差. 令 $\boldsymbol{\beta}^{(1)}$ 是 $(\boldsymbol{\Sigma}-\lambda_1\boldsymbol{I})\boldsymbol{\beta}=\boldsymbol{0}$ 的正规解. 则 $U_1=\boldsymbol{\beta}^{(1)'}\boldsymbol{X}$ 是有极大方差的正规线性组合. [如果 $\boldsymbol{\Sigma}-\lambda_1\boldsymbol{I}$ 的秩为 $p-1$, 则 $(\boldsymbol{\Sigma}-\lambda_1\boldsymbol{I})\boldsymbol{\beta}=0$ 只有一个解满足 $\boldsymbol{\beta}'\boldsymbol{\beta}=1$.]

现在我们来找所有线性组合中和 U_1 不相关的有极大方差的正规组合 $\boldsymbol{\beta}'\boldsymbol{X}$. 不相关意味着

$$0=\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X}U_1)=\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(1)})=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}, \tag{7}$$

因为 $\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=\lambda_1\boldsymbol{\beta}^{(1)}$. 因此 $\boldsymbol{\beta}'\boldsymbol{X}$ 和 U 正交, 无论是在统计意义 (不相关) 还是几何意义上 ($\boldsymbol{\beta}$ 和 $\boldsymbol{\beta}^{(1)}$ 的内积为零). (即当 $\lambda_1\neq 0$ 时, $\lambda_1\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$ 当且仅当 $\lambda_1\neq 0$ 时 $\boldsymbol{\beta}'\boldsymbol{\beta}^{(1)}=0$, 并且当 $\boldsymbol{\Sigma}\neq\mathbf{0}$ 时 $\lambda_1\neq 0$; $\boldsymbol{\Sigma}=\mathbf{0}$ 的情况是平凡的, 这里不再处理.)

现在我们要极大化

$$\phi_2=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}-\lambda(\boldsymbol{\beta}'\boldsymbol{\beta}-1)-2\nu_1\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}, \tag{8}$$

其中 λ 和 ν_1 是拉格朗日乘数. 偏导数的向量是

$$\frac{\partial\phi_2}{\partial\boldsymbol{\beta}}=2\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}-2\nu_1\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}, \tag{9}$$

我们令它为 $\mathbf{0}$. 在 (9) 式两边左乘 $\boldsymbol{\beta}'$ 并由 (7) 式我们得到

$$0=2\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\beta}-2\nu_1\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(1)}=-2\nu_1\lambda_1. \tag{10}$$

因此, $\nu_1=0$ 且 $\boldsymbol{\beta}$ 必须满足 (4) 式, 而 λ 必须满足 (5) 式. 令 $\lambda_{(2)}$ 为 $\lambda_1,\cdots,\lambda_p$ 中使得存在向量 $\boldsymbol{\beta}$ 满足 $(\boldsymbol{\Sigma}-\lambda_{(2)}\boldsymbol{I})\boldsymbol{\beta}=\mathbf{0}$, $\boldsymbol{\beta}'\boldsymbol{\beta}=1$ 和 (7) 式的最大者; 并称这个向量为 $\boldsymbol{\beta}^{(2)}$ 且相应的线性组合为 $U_2=\boldsymbol{\beta}^{(2)\prime}\boldsymbol{X}$. (后面将证明 $\lambda_{(2)}=\lambda_2$. 我们定义 $\lambda_{(1)}=\lambda_1$.)

继续这种方法, 在第 $(r+1)$ 步我们要找一个向量 $\boldsymbol{\beta}$, 它满足在和 $U_1,\cdots,U_r$ 不相关的所有正规线性组合中具有极大方差, 即满足

$$0=\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X}U_i)=\mathrm{E}(\boldsymbol{\beta}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\beta}^{(i)})=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)}=\lambda_{(i)}\boldsymbol{\beta}'\boldsymbol{\beta}^{(i)},\quad i=1,\cdots,r. \tag{11}$$

我们想极大化

$$\phi_{r+1}=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}-\lambda(\boldsymbol{\beta}'\boldsymbol{\beta}-1)-2\sum_{i=1}^{r}v_i\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)}, \tag{12}$$

其中 λ 和 $\nu_1,\cdots,\nu_r$ 是拉格朗日乘数. 偏导数的向量是

$$\frac{\partial\phi_{r+1}}{\partial\boldsymbol{\beta}}=2\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}-2\sum_{i=1}^{r}v_i\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)}, \tag{13}$$

且我们令它为 $\mathbf{0}$. 用 $\boldsymbol{\beta}^{(j)\prime}$ 左乘 (13) 式, 我们得到

$$0=2\boldsymbol{\beta}^{(j)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}-2\lambda\boldsymbol{\beta}^{(j)\prime}\boldsymbol{\beta}-2\nu_j\boldsymbol{\beta}^{(j)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)}. \tag{14}$$

如果 $\lambda_{(j)}\neq 0$, 则有 $-2\nu_j\lambda_{(j)}=0$ 且 $\nu_j=0$. 如果 $\lambda_{(j)}=0$, 则 $\boldsymbol{\Sigma}\boldsymbol{\beta}^{(j)}=\lambda_{(j)}\boldsymbol{\beta}^{(j)}=\mathbf{0}$ 且 (13) 式求和中的第 j 项就没有了. 因此 $\boldsymbol{\beta}$ 必须满足 (4) 式, 而 λ 必须满足 (5) 式.

令 $\lambda_{(r+1)}$ 为 $\lambda_1,\cdots,\lambda_p$ 中使得存在向量 $\boldsymbol{\beta}$ 满足 $(\boldsymbol{\Sigma}-\lambda_{(r+1)}\boldsymbol{I})\boldsymbol{\beta}=\mathbf{0},\boldsymbol{\beta}'\boldsymbol{\beta}=1$ 和 (11) 式的最大者, 称这个向量为 $\boldsymbol{\beta}^{(r+1)}$ 且相应的线性组合为 $U_{r+1}=\boldsymbol{\beta}^{(r+1)\prime}\boldsymbol{X}$. 如果 $\lambda_{(r+1)}=0$ 且 $\lambda_{(j)}=0, j\neq r+1$, 则 $\boldsymbol{\beta}^{(j)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(r+1)}=0$ 并不意味着 $\boldsymbol{\beta}^{(j)\prime}\boldsymbol{\beta}^{(r+1)}=0$. 但是 $\boldsymbol{\beta}^{(r+1)}$ 可以由 $\boldsymbol{\beta}^{(r+1)}$ 与所有满足 $\lambda_{(j)}$ 为 0 的 $\boldsymbol{\beta}^{(j)}$ 的线性组合代替, 这样新的 $\boldsymbol{\beta}^{(r+1)}$ 就和所有的 $\boldsymbol{\beta}^{(j)}, j=1,\cdots,r$ 正交. 继续这种方法直到第 $(m+1)$ 步我

们不能找到满足 $\beta'\beta=1$, (4) 式和 (11) 式的向量 β 为止. 由于 $\beta^{(1)},\cdots,\beta^{(m)}$ 必须线性独立所以 $m=p$ 或者 $m<p$.

现在我们将证明由不等式 $m<p$ 可以推出矛盾. 如果 $m<p$ 则存在 $p-m$ 个向量, 比如 $e_{m+1},\cdots,e_p$, 满足 $\beta^{(i)'}e_j=0$, $e_i'e_j=\delta_{ij}$. (这是由附录中的引理 A.4.2 得到的.) 令 $(e_{m+1},\cdots,e_p)=E$. 现在我们将证明存在一个 $(p-m)$ 元向量 c 和一个数 θ, 使得 $Ec=\sum c_ie_i$ 是当 $\lambda=\theta$ 时 (4) 式的根. 考虑 $|E'\Sigma E-\theta I|=0$ 的一个根和相应的满足 $E'\Sigma E=\theta c$ 的向量 c. 向量 ΣEc 和 $\beta^{(1)},\cdots,\beta^{(m)}$ 正交(由于 $\beta^{(i)'}\Sigma Ec=\lambda_{(i)}\beta^{(i)'}\sum c_je_j=\lambda_{(i)}\sum c_j\beta^{(i)'}e_j=0$), 因此是由 $e_{m+1},\cdots,e_p$ 张成的空间中的一个向量, 可以记为 Eg [其中 g 是一个 $(p-m)$ 元向量]. 在 $\Sigma Ec=Eg$ 两边左乘 E', 我们得到 $E'\Sigma Ec=E'Eg=g$. 因此 $g=\theta c$, 并且我们有 $\Sigma(Ec)=\theta(Ec)$. 则 $(Ec)'X$ 和 $\beta^{(j)'}X\ (j=1,\cdots,m)$ 不相关, 因此得到一个新的 $\beta^{(m+1)}$. 由于这和假设 $m<p$ 矛盾, 所以我们有 $m=p$.

令 $\mathbf{B}=(\beta^{(1)}\cdots\beta^{(p)})$ 且

$$\Lambda=\begin{pmatrix}\lambda_{(1)} & 0 & \cdots & 0\\ 0 & \lambda_{(2)} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_{(p)}\end{pmatrix}. \tag{15}$$

等式 $\Sigma\beta^{(r)}=\lambda_{(r)}\beta^{(r)}$ 可以记为矩阵形式, 如

$$\Sigma\mathbf{B}=\mathbf{B}\Lambda, \tag{16}$$

等式 $\beta^{(r)'}\beta^{(r)}=1$ 和 $\beta^{(r)'}\beta^{(s)}=0$, $r\neq s$, 可以记为

$$\mathbf{B}'\mathbf{B}=I. \tag{17}$$

从 (16) 式和 (17) 式我们得到

$$\mathbf{B}'\Sigma\mathbf{B}=\Lambda. \tag{18}$$

从

$$\begin{aligned}|\Sigma-\lambda I| &=|\mathbf{B}'|\cdot|\Sigma-\lambda I|\cdot|\mathbf{B}|\\ &=|\mathbf{B}'\Sigma\mathbf{B}-\lambda\mathbf{B}'\mathbf{B}|=|\Lambda-\lambda I|\\ &=\Pi(\lambda_{(i)}-\lambda)\end{aligned} \tag{19}$$

我们知道 (19) 式的根是 Λ 的对角元素, 即 $\lambda_{(1)}=\lambda_1$, $\lambda_{(2)}=\lambda_2,\cdots,\lambda_{(p)}=\lambda_p$.

我们已经证明了下面的定理.

定理 11.2.1 令 p 元随机向量 X 满足 $\mathrm{E}(X)=\mathbf{0}$ 且 $\mathrm{E}(XX')=\Sigma$. 则存在正交线性变换

$$U=\mathbf{B}'X \tag{20}$$

使得 U 的协方差阵是 $\mathrm{E}(UU')=\Lambda$ 且

$$\mathbf{\Lambda}=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_p\end{pmatrix}, \tag{21}$$

其中 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$ 是 (5) 式的根. $\mathbf{B}$的第 r 列 $\boldsymbol{\beta}^{(r)}$ 满足 $(\mathbf{\Sigma}-\lambda_r \boldsymbol{I})\boldsymbol{\beta}^{(r)}=\mathbf{0}$. $\boldsymbol{U}$ 的第 r 个分量 $U_r=\boldsymbol{\beta}^{(r)'}\boldsymbol{X}$ 在所有和 $U_1,\cdots,U_{r-1}$ 不相关的正规线性组合中具有极大方差.

向量 $\boldsymbol{U}$ 定义为 $\boldsymbol{X}$ 的主成分向量. 在附录 A 中我们已经证明了定理 A.2.2, 它对于半正定的 $\boldsymbol{B}$ 成立, 事实上, 对于任意对称的 $\boldsymbol{B}$, 这个证明都成立. 需要注意的是, 一旦变换到 $U_1,\cdots,U_p$, 显然 U_1 是具有极大方差的正规线性组合, 因为如果 $U^*=\sum c_i U_i$, 其中 $\sum c_i^2=1(U^*$ 也是 $\boldsymbol{X}$ 的正规线性组合), 则 $\operatorname{Var}(U^*)=\sum c_i^2\lambda_i=\lambda_1+\sum_{i=2}^p c_i^2(\lambda_i-\lambda_1)$ (因为 $c_1^2=1-\sum_2^p c_i^2$), 显然当 $c_i^2=0, i=2,\cdots,p$ 时有极大值. 类似地, U_2 是和 U_1 不相关的正规线性组合且具有极大方差 ($U^*=\sum c_i U_i$ 和 U_1 不相关意味着 $c_1=0$); 依次地, $U_3,\cdots,U_p$ 的极大性质也可以证明.

我们也可以得到一些其他的结论.

引理 11.2.1 *假设 $\lambda_{r+1}=\cdots=\lambda_{r+m}=\nu$ (即 ν 是一个 m 重根), 则 $\mathbf{\Sigma}-\nu\boldsymbol{I}$ 的秩为 $p-m$. 而且 $\mathbf{B}^*=(\boldsymbol{\beta}^{(r+1)}\cdots\boldsymbol{\beta}^{(r+m)})$ 除了重数外是由一个正交矩阵唯一确定的.*

证明 从定理的推导我们有 $(\mathbf{\Sigma}-\nu\boldsymbol{I})\boldsymbol{\beta}^{(i)}=\mathbf{0}, i=r+1,\cdots,r+m$, 即 $\boldsymbol{\beta}^{(r+1)},\cdots,\boldsymbol{\beta}^{(r+m)}$ 是 $(\mathbf{\Sigma}-\nu\boldsymbol{I})\boldsymbol{\beta}=\mathbf{0}$ 的 m 个线性独立的解. 为了证明不存在另外的线性独立的解, 取 $\sum_{i=1}^p x_i\boldsymbol{\beta}^{(i)}$, 其中 x_i 是标量. 如果它是解, 则有 $\nu\sum x_i\boldsymbol{\beta}^{(i)}=\mathbf{\Sigma}(\sum x_i\boldsymbol{\beta}^{(i)})=\sum x_i\mathbf{\Sigma}\boldsymbol{\beta}^{(i)}=\sum x_i\lambda_i\boldsymbol{\beta}^{(i)}$. 因为 $\nu x_i=\lambda_i x_i$, 所以除非 $i=r+1,\cdots,r+m$ 否则一定有 $x_i=0$. 因此秩为 $p-m$.

如果 $\mathbf{B}^*$ 是 $(\mathbf{\Sigma}-\nu\boldsymbol{I})\boldsymbol{\beta}=\mathbf{0}$ 的一组解, 则其他任意一组解一定是这组解的线性组合, 即是 $\mathbf{B}^*\boldsymbol{A}$, 其中 $\boldsymbol{A}$ 非奇异. 但是, 把正交的条件 $\mathbf{B}^{*'}\mathbf{B}^*=\boldsymbol{I}$ 应用到线性组合上得 $\boldsymbol{I}=(\mathbf{B}^*\boldsymbol{A})'(\mathbf{B}^*\boldsymbol{A})=\boldsymbol{A}'\mathbf{B}^{*'}\mathbf{B}^*\boldsymbol{A}=\boldsymbol{A}'\boldsymbol{A}$, 因此 $\boldsymbol{A}$ 一定是正交的. ■

定理 11.2.2 *随机向量 $\boldsymbol{X}$ 的正交变换 $\boldsymbol{V}=\boldsymbol{C}\boldsymbol{X}$ 保持广义方差和分量的方差和不变.*

证明 令 $\mathrm{E}(\boldsymbol{X})=\mathbf{0}$ 且 $\mathrm{E}(\boldsymbol{X}\boldsymbol{X}')=\mathbf{\Sigma}$. 则 $\mathrm{E}(\boldsymbol{V})=\mathbf{0}$ 且 $\mathrm{E}(\boldsymbol{V}\boldsymbol{V}')=\boldsymbol{C}\mathbf{\Sigma}\boldsymbol{C}'$. $\boldsymbol{V}$ 的广义方差是

$$|\boldsymbol{C}\mathbf{\Sigma}\boldsymbol{C}'|=|\boldsymbol{C}|\cdot|\mathbf{\Sigma}|\cdot|\boldsymbol{C}'|=|\mathbf{\Sigma}|\cdot|\boldsymbol{C}\boldsymbol{C}'|=|\mathbf{\Sigma}|, \tag{22}$$

这也是 $\boldsymbol{X}$ 的广义方差. $\boldsymbol{V}$ 的分量的方差的和是

$$\sum \mathrm{E}(V_i^2)=\operatorname{tr}(\boldsymbol{C}\mathbf{\Sigma}\boldsymbol{C}')=\operatorname{tr}(\mathbf{\Sigma}\boldsymbol{C}'\boldsymbol{C})=\operatorname{tr}(\mathbf{\Sigma}\boldsymbol{I})=\operatorname{tr}\mathbf{\Sigma}=\sum \mathrm{E}(X_i^2). \quad ■ \tag{23}$$

推论 11.2.2 *主成分向量的广义方差是原来向量的广义方差, 且主成分的方差和是原来变量的方差和.*

得到上面理论的另一种方法是根据均值向量为 0 和协方差阵为 $\boldsymbol{\Sigma}$(非奇异) 的正态分布的常密度曲面. 密度是

$$\frac{1}{(2\pi)^{\frac{1}{2}p}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}, \tag{24}$$

且常密度曲面是椭球面

$$\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = C. \tag{25}$$

这个椭球面的主轴定义为从 $-\boldsymbol{y}$ 到 $\boldsymbol{y}$ 的一条线, 其中 $\boldsymbol{y}$ 是平方距离 $\boldsymbol{x}'\boldsymbol{x}$ 有平稳点的椭球上的一点. 利用拉格朗日乘数的方法, 我们通过考虑下式来确定平稳点

$$\psi = \boldsymbol{x}'\boldsymbol{x} - \lambda\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}, \tag{26}$$

其中 λ 是拉格朗日乘数. 我们对 ψ 关于 $\boldsymbol{x}$ 求导, 且令导数为 0, 得到

$$\frac{\partial\psi}{\partial\boldsymbol{x}} = 2\boldsymbol{x} - 2\lambda\boldsymbol{\Sigma}^{-1}\boldsymbol{x} = \boldsymbol{0}. \tag{27}$$

或者

$$\boldsymbol{x} = \lambda\boldsymbol{\Sigma}^{-1}\boldsymbol{x}. \tag{28}$$

乘以 $\boldsymbol{\Sigma}$ 得到

$$\boldsymbol{\Sigma}\boldsymbol{x} = \lambda\boldsymbol{x}. \tag{29}$$

这个等式和 (4) 相同并且可以进行同样的代数运算. 因此向量 $\boldsymbol{\beta}^{(1)},\cdots,\boldsymbol{\beta}^{(p)}$ 给出了椭球面的主轴. 变换 $\boldsymbol{u} = \boldsymbol{\beta}'\boldsymbol{x}$ 是坐标轴的一个旋转, 使得新的坐标轴在椭球面的主轴的方向. 在新的坐标轴下椭球面是

$$\boldsymbol{u}'\boldsymbol{\Lambda}^{-1}\boldsymbol{u} = \sum\frac{u_i^2}{\lambda_i} = C. \tag{30}$$

因此第 i 个主轴的长度是 $2\sqrt{\lambda_i C}$.

第三种方法是利用最近拟合的平面 [Pearson(1901)], 也可以得到同样的结果. 考虑通过原点的一个平面 $\boldsymbol{\alpha}'\boldsymbol{x} = 0$, 其中 $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. 点 $\boldsymbol{x}$ 到平面的距离为 $\boldsymbol{\alpha}'\boldsymbol{x}$. 让我们找到一个平面的系数使得一个随机点 $\boldsymbol{X}$ 到平面的期望距离平方是极小值, 其中 $\mathrm{E}(\boldsymbol{X}) = 0$ 且 $\mathrm{E}(\boldsymbol{X}\boldsymbol{X}') = \boldsymbol{\Sigma}$. 因此我们想要极小化 $\mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X})^2 = \mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\alpha}) = \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}$, 在限制条件 $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$ 下. 和第一种方法对比我们很快就知道解为 $\boldsymbol{\alpha} = \boldsymbol{\beta}^{(p)}$.

当 $\boldsymbol{X}$ 的所有分量都在同一个单位下测得时, 对于主分量的分析最合适的. 如果它们的度量单位不一样, 关于 $\boldsymbol{\beta}'\boldsymbol{\beta}$ 极大化 $\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ 的合理性是值得怀疑的. 事实上, 这个分析将依赖于不同的测量单位. 假设 $\boldsymbol{\Delta}$ 是一个对角矩阵, 且令 $\boldsymbol{Y} = \boldsymbol{\Delta}\boldsymbol{X}$. 例如, $\boldsymbol{X}$ 的一个分量可能是用英寸来度量的而相应的 $\boldsymbol{Y}$ 的分量是用英尺来度量的; $\boldsymbol{X}$ 的另一个分量可能以磅为单位而相应的 $\boldsymbol{Y}$ 的分量则以盎司为单位. $\boldsymbol{Y}$ 的协方

差阵是 $\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}') = \mathrm{E}(\boldsymbol{\Delta}\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\Delta}) = \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta} = \boldsymbol{\Psi}$. 关于 $\boldsymbol{Y}$ 的主成分分析包括关于 $\gamma'\gamma$ 极大化 $\mathrm{E}(\gamma'\gamma)^2 = \gamma'\boldsymbol{\Psi}\gamma$, 且得到等式 $\mathbf{0} = (\boldsymbol{\Psi} - \nu\boldsymbol{I})\gamma = (\boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta} - \nu\boldsymbol{I})\gamma$, 其中 ν 满足 $|\boldsymbol{\Psi} - \nu\boldsymbol{I}| = 0$. 左乘 $\boldsymbol{\Delta}^{-1}$ 得到

$$\mathbf{0} = (\boldsymbol{\Sigma} - \nu\boldsymbol{\Delta}^{-2})(\boldsymbol{\Delta}\gamma). \tag{31}$$

令 $\boldsymbol{\Delta}\gamma = \boldsymbol{\alpha}$, 即 $\gamma'\boldsymbol{Y} = \gamma'\boldsymbol{\Delta}\boldsymbol{X} = \boldsymbol{\alpha}'\boldsymbol{X}$. 那么 (31) 式可以由关于 $\boldsymbol{\alpha}'\boldsymbol{\Delta}^{-2}\boldsymbol{\alpha}$ 极大化 $E(\boldsymbol{\alpha}'\boldsymbol{X})^2 = \boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}$ 来得到. 最后一个二次型是一个平方的加权和的形式, 其中权重是 $\boldsymbol{\Delta}^{-2}$ 的对角元素.

需要注意的是如果 $\boldsymbol{\Delta}^{-2}$ 是这种形式的矩阵,

$$\boldsymbol{\Delta}^{-2} = \begin{pmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{pmatrix}, \tag{32}$$

则 $\boldsymbol{\Psi}$ 是相关阵.

11.3 主成分和它们的方差的极大似然估计

在主成分分析中, 一个基本的统计推断问题是估计向量 $\boldsymbol{\beta}^{(1)}, \cdots, \boldsymbol{\beta}^{(p)}$ 和标量 $\lambda_1, \cdots, \lambda_p$. 我们应用前面几节的代数知识去估计相关阵.

定理 11.3.1 令 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是总体 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 中的 $N (> p)$ 个观测, 其中 $\boldsymbol{\Sigma}$ 是有 p 个不同特征根的矩阵. 则定理 11.2.1 中定义的 $\lambda_1, \cdots, \lambda_p$ 和 $\boldsymbol{\beta}^{(1)}, \cdots, \boldsymbol{\beta}^{(p)}$ 的极大似然估计量分别是

$$|\hat{\boldsymbol{\Sigma}} - k\boldsymbol{I}| = 0 \tag{1}$$

的根 $k_1 \geqslant \cdots \geqslant k_p$ 和满足

$$(\hat{\boldsymbol{\Sigma}} - k_i\boldsymbol{I})\boldsymbol{b}^{(i)} = \mathbf{0}, \tag{2}$$

$$\boldsymbol{b}^{(i)\prime}\boldsymbol{b}^{(i)} = 1 \tag{3}$$

的相应向量 $\boldsymbol{b}^{(1)}, \cdots, \boldsymbol{b}^{(p)}$, 其中 $\hat{\boldsymbol{\Sigma}}$ 是 $\boldsymbol{\Sigma}$ 的极大似然估计.

证明 当 $|\boldsymbol{\Sigma} - \lambda\boldsymbol{I}| = 0$ 的根不同时, 每一个向量 $\boldsymbol{\beta}^{(i)}$ 都唯一确定除非用 $-\boldsymbol{\beta}^{(i)}$ 来代替 $\boldsymbol{\beta}^{(i)}$. 如果我们要求 $\boldsymbol{\beta}^{(i)}$ 的第一个非零分量为正的, 则 $\boldsymbol{\beta}^{(i)}$ 是唯一确定的, 且 $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{B}$ 是 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ 的单值函数. 由推论 3.2.1 知, $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{B}$ 的极大似然估计是 $\hat{\mu}, \hat{\boldsymbol{\Sigma}}$ 的相同的函数. 这个函数由 (1), (2) 和 (3) 确定且也有 $\boldsymbol{b}^{(i)}$ 的第一个非零分量为正的相应限制. [可以证明, 如果 $|\boldsymbol{\Sigma}| \neq 0$, 则 (1) 式的根以概率 1 是不同的, 因为 $\hat{\boldsymbol{\Sigma}}$ 有重数大于 1 的根的条件确定了维数小于 $\frac{1}{2}p(p+1)$ 的 $\hat{\boldsymbol{\Sigma}}$ 的空间的一个区域, 参见 Okamoto(1973).] 从 11.2 节的 (18) 式我们知道

$$\boldsymbol{\Sigma}=\boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}'=\sum\lambda_i\boldsymbol{\beta}^{(i)}\boldsymbol{\beta}^{(i)\prime},\tag{4}$$

且由同样的代数运算知

$$\hat{\boldsymbol{\Sigma}}=\sum k_i\boldsymbol{b}^{(i)}\boldsymbol{b}^{(i)\prime}.\tag{5}$$

显然把 $\boldsymbol{b}^{(i)}$ 换为 $-\boldsymbol{b}^{(i)}$ 不能改变 $\sum k_i\boldsymbol{b}^{(i)}\boldsymbol{b}^{(i)\prime}$. 因为似然函数依赖于 $\hat{\boldsymbol{\Sigma}}$(参见 3.2 节), 似然函数的极大值在 (2) 式和 (3) 式的任意解处达到. ■

明确假设 $\boldsymbol{\Sigma}$ 的根是有任意重数的是可能的. 如果这些重数不都是 1, 则极大似然估计不是定理 11.3.1 中定义的. [参见 Anderson(1963a).] 例如假设我们知道等式 $|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|=0$ 有一个重数为 p 的根. 令这个根为 λ_1. 则由推论 11.2.1 知, $\boldsymbol{\Sigma}-\lambda_1\boldsymbol{I}$ 的秩为 0, 即 $\boldsymbol{\Sigma}-\lambda_1\boldsymbol{I}=\boldsymbol{0}$ 或者 $\boldsymbol{\Sigma}=\lambda_1\boldsymbol{I}$. 如果 $\boldsymbol{X}$ 的分布服从 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})=N(\boldsymbol{\mu},\lambda_1\boldsymbol{I})$, 则 $\boldsymbol{X}$ 的分量独立地分布且有方差 λ_1. 因此 λ_1 的极大似然估计量是

$$\hat{\lambda_1}=\frac{1}{pN}\sum_{i=1}^{p}\sum_{\alpha=1}^{N}(x_{i\alpha}-\bar{x}_i)^2,\tag{6}$$

$\hat{\boldsymbol{\Sigma}}=\hat{\lambda_1}\boldsymbol{I}$, 而 $\hat{\boldsymbol{\beta}}$ 可以是任意的正交矩阵. 需要指出的是, 10.7 节中我们考虑过假设检验 $\boldsymbol{\Sigma}=\lambda_1\boldsymbol{I}$ (λ_1 没有给出), 即假设 $\boldsymbol{\Sigma}$ 有一个重数为 p 的特征根.

在许多主成分分析的应用中一般假设 $\boldsymbol{\Sigma}$ 的根是不同的. 特别指出的是, 在这种方法的一些应用中代数也会用于相关矩阵而不是协方差阵. 一般地这会得出不同的根和向量.

11.4 主成分的极大似然估计的计算

计算一个矩阵 $\boldsymbol{\Sigma}$ 或 $\hat{\boldsymbol{\Sigma}}$ 的特征根和特征向量 (主成分) 的方法有很多种. 下面我们给出其中的几种.

对于小 p, 一种方法就是展开行列式方程

$$0=|\boldsymbol{\Sigma}-\lambda\boldsymbol{I}|\tag{1}$$

并且解出得到的 λ 的 p 次方程的根 $\lambda_1>\lambda_2>\cdots>\lambda_p$(例如, 利用牛顿法或割线法). 则 $\boldsymbol{\Sigma}-\lambda_i\boldsymbol{I}$ 的秩为 $p-1$, 且 $(\boldsymbol{\Sigma}-\lambda_i\boldsymbol{I})\boldsymbol{\beta}^{(i)}=\boldsymbol{0}$ 的解可以通过把 $\boldsymbol{\beta}$ 取为 $\boldsymbol{\Sigma}-\lambda_i\boldsymbol{I}$ 的第一列 (或者任意固定的一列) 和第 j 行的元素的余子式得到.

第二种方法利用特征根和特征向量的方程来迭代

$$\boldsymbol{\Sigma}\boldsymbol{x}=\lambda\boldsymbol{x},\tag{2}$$

其中我们已经把方程写成总体的方程. 令 $\boldsymbol{x}_{(0)}$ 是不和第一个特征向量正交的任意向量, 并且定义

$$\boldsymbol{x}_{(i)}=\boldsymbol{\Sigma}\boldsymbol{y}_{(i-1)},\quad \boldsymbol{y}_{(i)}=\frac{1}{\sqrt{\boldsymbol{x}_{(i)}'\boldsymbol{x}_{(i)}}}\boldsymbol{x}_{(i)},\quad i=0,1,2,\cdots.\tag{3}$$

可以证明 (习题 11.12)

$$\lim_{i\to\infty} \boldsymbol{y}_{(i)} = \pm\boldsymbol{\beta}^{(1)}, \quad \lim_{i\to\infty} \boldsymbol{x}'_{(i)}\boldsymbol{x}_{(i)} = \lambda_1^2. \tag{4}$$

收敛速度依赖于比值 λ_2/λ_1; 比值越接近于 1, 收敛速度越慢.

为了找到第二个根和向量, 定义

$$\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma} - \lambda_1\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(1)\prime}. \tag{5}$$

则如果 $i \neq 1$,

$$\boldsymbol{\Sigma}_2\boldsymbol{\beta}^{(i)} = \boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} - \lambda_1\,\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\beta}^{(i)} = \boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)} = \lambda_i\boldsymbol{\beta}^{(i)}, \tag{6}$$

$$\boldsymbol{\Sigma}_2\boldsymbol{\beta}^{(1)} = \boldsymbol{0}. \tag{7}$$

因此 λ_2 是 $\boldsymbol{\Sigma}_2$ 的最大的根而 $\boldsymbol{\beta}^{(2)}$ 是相应的向量. 这个迭代过程现在应用于 $\boldsymbol{\Sigma}_2$, 且用来寻找 λ_2 和 $\boldsymbol{\beta}^{(2)}$. 定义 $\boldsymbol{\Sigma}_3 = \boldsymbol{\Sigma}_2 - \lambda_2\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\prime}$, 我们可以找到 λ_3 和 $\boldsymbol{\beta}^{(3)}$, 等等.

迭代方法比较费力, 可以有几种方法来改进. 一种是在迭代之前先把 $\boldsymbol{\Sigma}$ 提高到幂. 因此我们可以利用 $\boldsymbol{\Sigma}^2$, 定义

$$\boldsymbol{x}_{(i)} = \boldsymbol{\Sigma}^2\boldsymbol{y}_{(i-1)}, \quad \boldsymbol{y}_{(i)} = \frac{\boldsymbol{x}_{(i)}}{\sqrt{\boldsymbol{x}'_{(i)}\boldsymbol{x}_{(i)}}}, \quad i = 0, 1, 2, \cdots. \tag{8}$$

这种方法的收敛速度比利用 (3) 式快一倍. 利用 $\boldsymbol{\Sigma}^4 = \boldsymbol{\Sigma}^2\boldsymbol{\Sigma}^2$ 可以把收敛速度提高四倍, 依次类推. 需要注意的是, 由于 $\boldsymbol{\Sigma}^2$ 是对称的, 所以只能有 $p(p+1)/2$ 个元素.

但是有效的计算还可以利用其他方法. 一种方法是 QR 或 QL 算法. 令 $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}$. 利用 $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i\boldsymbol{L}_i$ 和 $\boldsymbol{\Sigma}_{i+1} = \boldsymbol{L}_i\boldsymbol{Q}_i\,(= \boldsymbol{Q}_i'\boldsymbol{\Sigma}_i\boldsymbol{Q}_i)$ 依次定义正交矩阵 $\boldsymbol{Q}_i$ 和下三角矩阵 $\boldsymbol{L}_i$, $i = 1, 2, \cdots$.(格拉姆-施密特正交化是寻找 $\boldsymbol{Q}_i$ 和 $\boldsymbol{L}_i$ 的一种方法; QR 方法把下三角矩阵 $\boldsymbol{L}$ 替换为上三角矩阵 $\boldsymbol{R}$.) 如果 $\boldsymbol{\Sigma}$ 的特征根不同, 则 $\lim_{i\to\infty}\boldsymbol{\Sigma}_{i+1} = \boldsymbol{\Lambda}^*$, 其中 $\boldsymbol{\Lambda}^*$ 是对角矩阵, 对角元素是按从大到小顺序排列的根. 特征向量是 $\lim_{i\to\infty}\boldsymbol{Q}_i'\boldsymbol{Q}_{i-1}'\cdots\boldsymbol{Q}_1'$ (依次计算出来) 的列.

更有效的算法 (对于对称的 $\boldsymbol{\Sigma}$) 是利用一串 Householder 变换把 $\boldsymbol{\Sigma}$ 变为对角形式. 一个 Householder 矩阵是 $\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{\alpha}\boldsymbol{\alpha}'$, 其中 $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. 这样的矩阵是正交且对称的. 对称矩阵 $\boldsymbol{\Sigma}$ 的 Householder 变换是 $\boldsymbol{H\Sigma H}$. 它是对称的且和 $\boldsymbol{\Sigma}$ 有相同的特征根, 它的特征向量是 $\boldsymbol{H}$ 乘以 $\boldsymbol{\Sigma}$ 的特征向量.

三对角矩阵是除了主对角和两个次对角线外, 其余元素都是 0 的矩阵. $p-2$ 个 Householder 变换后对称的 $\boldsymbol{\Sigma}$ 变为三对角形式. (第一次变换把 $\boldsymbol{H\Sigma H}$ 的第一列和第一行的最后 $p-2$ 个元素变为 0, 依次类推. 见习题 11.13.)

我们把 QL 方法用于三对角矩阵. 在第 i 步令三对角矩阵为 $\boldsymbol{T}_0^{(i)}$; 令 $\boldsymbol{P}_j^{(i)}$ 为块对角矩阵 (吉文斯矩阵)

$$P_j^{(i)} = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & \cos\theta_j & -\sin\theta_j & 0 \\ 0 & \sin\theta_j & \cos\theta_j & 0 \\ 0 & 0 & 0 & I \end{bmatrix}, \tag{9}$$

其中 $\cos\theta_j$ 是第 j 个和第 $j+1$ 个对角元素; 令 $\boldsymbol{T}_j^{(i)} = \boldsymbol{P}_{p-j}^{(i)}\boldsymbol{T}_{j-1}^{(i)}$, $j=1,\cdots,p-1$. 这里 θ_j 用来使得 $\boldsymbol{T}_j$ 的位置 $j,j+1$ 上的元素为 0. 则 $\boldsymbol{P}^{(i)} = \boldsymbol{P}_1^{(i)}\boldsymbol{P}_2^{(i)}\cdots\boldsymbol{P}_{p-1}^{(i)}$ 是正交的且 $\boldsymbol{P}^{(i)}\boldsymbol{T}_0^{(i)} = \boldsymbol{R}^{(i)}$ 是下三角的. 则 $\boldsymbol{T}_0^{(i+1)} = \boldsymbol{R}^{(i)}\boldsymbol{P}^{(i)\prime}(=\boldsymbol{P}^{(i)}\boldsymbol{T}_0^{(i)}\boldsymbol{P}^{(i)\prime})$ 是对称的和三对角的. 它收敛于 $\boldsymbol{\Lambda}^*$(如果根都不相同). 更多的细节可以参考 Wilkinson and Reinsch(1971) 的第 II/2 章和第 II/3 章, Wilkinson(1965) 的第 5 章, Golub and Van Loan(1989) 的第 5, 7, 8 章. 一系列单边 Householder 变换 ($\boldsymbol{H\Sigma}$) 可以把 $\boldsymbol{\Sigma}$ 变为 $\boldsymbol{R}$(上三角), 从而得到 QR 分解.

11.5 例　子

在表 3.4 中我们列出了鸢尾品种的三个观测样本 [Fisher(1936)], 作为主成分分析的一个例子我们利用其中一个样本, 即**变色鸢尾**. 有 50 个观测 ($N=50, n=N-1=49$). 每一个观测包括一株植物的四个测量: x_1 表示花萼长度, x_2 表示花萼宽度, x_3 表示花瓣长度, x_4 表示花瓣宽度. 观测的离差平方和是

$$\boldsymbol{A} = \sum_{\alpha=1}^{50}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = \begin{pmatrix} 13.0552 & 4.1740 & 8.9620 & 2.7332 \\ 4.1740 & 4.8250 & 4.0500 & 2.0190 \\ 8.9620 & 4.0500 & 10.8200 & 3.5820 \\ 2.7332 & 2.0190 & 3.5820 & 1.9162 \end{pmatrix}, \tag{1}$$

且 $\boldsymbol{\Sigma}$ 的估计是

$$S = \frac{1}{49}\boldsymbol{A} = \begin{pmatrix} 0.266\,433 & 0.085\,184 & 0.182\,899 & 0.055\,780 \\ 0.085\,184 & 0.098\,469 & 0.082\,653 & 0.041\,204 \\ 0.182\,899 & 0.082\,653 & 0.220\,816 & 0.073\,102 \\ 0.055\,780 & 0.041\,204 & 0.073\,102 & 0.039\,106 \end{pmatrix}. \tag{2}$$

我们利用迭代方法来找第一个主成分, 依次算出 $\boldsymbol{z}^{(j)} = \boldsymbol{S}\boldsymbol{z}^{(j-1)}$. 作为一个初始近似, 我们设 $\boldsymbol{z}^{(0)\prime} = (1,0,1,0)$. 没有必要在每次迭代时都正规化向量; 但是为了对比前后的两个向量, 我们计算 $z_i^{(j)}/z_i^{(j-1)} = r_i^{(j)}$, 其中每一个都是 $\boldsymbol{S}$ 的最大根 l_1 的近似. 在第七次迭代后, $r_i^{(7)}$ 的误差小于 0.000 02(第五个显著数字). 正规化这个向量, 且把 $\boldsymbol{S}$ 应用于这个正规向量上. 比值 $r_i^{(8)}$ 的误差小于 0.000 002; l_1 的值 (几乎精确到第六位) 是 $l_1 = 0.487\,875$. 正规化的第八个迭代向量就是 $\boldsymbol{\beta}^{(1)}$ 的估计, 即

$$b^{(1)} = \begin{pmatrix} 0.686\,724\,4 \\ 0.305\,346\,3 \\ 0.623\,662\,8 \\ 0.214\,983\,7 \end{pmatrix}. \tag{3}$$

这个向量和第七次迭代的正规向量之差小于 0.000 001. 需要指出的是, l_1 和 $\boldsymbol{b}^{(1)}$ 必须比 l_2 和 $\boldsymbol{b}^{(2)}$ 计算得更精确, 依次类推. $\boldsymbol{S}$ 的迹是 0.624 824, 这是所有根的和. 因此 l_1 比其他根的和的三倍都要大.

接下来我们计算

$$\begin{aligned} \boldsymbol{S}_2 &= \boldsymbol{S} - l_1\boldsymbol{b}^{(1)}\boldsymbol{b}^{(1)\prime} \\ &= \begin{pmatrix} 0.036\,355\,9 & -0.017\,117\,9 & -0.026\,050\,2 & -0.016\,247\,2 \\ -0.017\,117\,9 & 0.052\,981\,3 & -0.010\,254\,6 & 0.009\,177\,7 \\ -0.026\,050\,2 & -0.010\,254\,6 & 0.031\,054\,4 & 0.007\,689\,0 \\ -0.016\,247\,2 & 0.009\,177\,7 & 0.007\,689\,0 & 0.016\,557\,4 \end{pmatrix}, \end{aligned} \tag{4}$$

且利用 $\boldsymbol{z}^{(0)\prime} = (0,1,0,0)$ 迭代计算 $\boldsymbol{z}^{(j)} = \boldsymbol{S}_2\boldsymbol{z}^{(j-1)}$. (实际计算中 $\boldsymbol{S}_2$ 先乘以 10, 并且第一行和第一列都乘以 −1.) 在这种情况下迭代的速度将减慢. 我们将会看到, l_2 和 l_3 的比值近似为 1.32. 在最后一次迭代时, 比值的误差小于 0.000 04. 我们得到 $l_2 = 0.072\,382\,8$ 且

$$b^{(2)} = \begin{pmatrix} -0.669\,033 \\ 0.567\,484 \\ 0.343\,309 \\ 0.335\,307 \end{pmatrix}. \tag{5}$$

第三个主成分从 $\boldsymbol{S}_3 = \boldsymbol{S}_2 - l_2\boldsymbol{b}^{(2)}\boldsymbol{b}^{(2)\prime}$ 中得到, 第四个从 $\boldsymbol{S}_4 = \boldsymbol{S}_3 - l_3\boldsymbol{b}^{(3)}\boldsymbol{b}^{(3)\prime}$ 中得到.

结果总结如下:

$$(l_1, l_2, l_3, l_4) = (0.4879, 0.0724, 0.0548, 0.0098), \tag{6}$$

$$\boldsymbol{B} = \begin{pmatrix} 0.6867 & -0.6690 & -0.2651 & 0.1023 \\ 0.3053 & 0.5675 & -0.7296 & -0.2289 \\ 0.6237 & 0.3433 & 0.6272 & -0.3160 \\ 0.2150 & 0.3353 & 0.0637 & 0.9150 \end{pmatrix}. \tag{7}$$

这四个根的和是 $\sum_{i=1}^4 l_i = 0.6249$, 而样本协方差阵的迹是 $\operatorname{tr}\boldsymbol{S} = 0.624\,824$. 四个值中第一个占了总方差的 78%; 最后一个占了 1% 多一点. 事实上, $0.7\boldsymbol{x}_1 + 0.3\boldsymbol{x}_2 + 0.6\boldsymbol{x}_3 + 0.2\boldsymbol{x}_4$(第一个主成分的近似) 的方差是 0.478, 差不多是总方差的 77%. 如果有人对研究方差从而对 (x_1, x_2, x_3, x_4) 的方差感兴趣, 那么他可以寻找 $0.7x_1 +$

$0.3x_2+0.6x_3+0.2x_4$ 的方差基础上的方差. 在探索性研究中忽略掉 (x_1,x_2,x_3,x_4) 的其他方差是不重要的.

11.6 统计推断

11.6.1 渐近分布

如果总体协方差阵是 $\boldsymbol{I}$ 或和 $\boldsymbol{I}$ 成比例, 即在总体的根都相等时, 在 13.3 节中我们将得到样本特征根和特征向量的精确分布. 当总体的根不都相等时, 根和向量的精确分布包含一个带多项式的倍增无穷级数, 这已经超出了本书的范围. [参见 Muirhead(1982).] 我们得到了当总体的根都不相同 (定理 13.5.1) 和有一个根是重根 (定理 13.5.2) 时根和向量的渐近分布. 除非特别指出, 通常可以认为总体的根是不同的, 所以我们这里推导定理 13.5.1.

和以前一样, 令 $\boldsymbol{\Sigma}$ 的特征根为 $\lambda_1>\cdots>\lambda_p$ 且相应的特征向量为 $\boldsymbol{\beta}^{(1)},\cdots,\boldsymbol{\beta}^{(p)}$, 它们都是正规化的, 故 $\boldsymbol{\beta}^{(i)\prime}\boldsymbol{\beta}^{(i)}=1$ 且满足 $\beta_{1i}\geqslant 0,\ i=1,\cdots,p$. 令 $\boldsymbol{S}$ 的特征根和特征向量为 $l_1>\cdots>l_p$ 和 $\boldsymbol{b}^{(1)},\cdots,\boldsymbol{b}^{(p)}$, 它们都是正规化的, 故 $\boldsymbol{b}^{(i)\prime}\boldsymbol{b}^{(i)}=1$, 且满足 $b_{1i}\geqslant 0,\ i=1,\cdots,p$. 令 $d_i=\sqrt{n}(l_i-\lambda_i)$ 且 $\boldsymbol{g}^{(i)}=\sqrt{n}(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)}),\ i=1,\cdots,p$. 则在极限正态分布中集合 $d_1,\cdots,d_p$ 和 $\boldsymbol{g}^{(1)},\cdots,\boldsymbol{g}^{(p)}$ 是独立的且 $d_1,\cdots,d_p$ 是相互独立的. 元素 d_i 有极限分布 $N(0,2\lambda_i^2)$. 在极限分布中 $\boldsymbol{g}^{(1)},\cdots,\boldsymbol{g}^{(p)}$ 的协方差是

$$\mathrm{Cov}(\boldsymbol{g}^{(i)})=\sum_{\substack{k=1\\k\neq i}}^{p}\frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\boldsymbol{\beta}^{(k)}\boldsymbol{\beta}^{(k)\prime},\tag{1}$$

$$\mathrm{Cov}(\boldsymbol{g}^{(i)},\boldsymbol{g}^{(j)})=-\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\boldsymbol{\beta}^{(j)}\boldsymbol{\beta}^{(i)\prime},\quad i\neq j.\tag{2}$$

见定理 13.5.1.

为了对有次序的单个根作出推断. 我们把 l_i 近似看作均值为 λ_i 方差为 $2\lambda_i^2/n$ 的正态分布. 由于 l_i 是 λ_i 的相容估计, 所以

$$\sqrt{n}\frac{l_i-\lambda_i}{\sqrt{2}l_i}\tag{3}$$

的极限分布是 $N(0,1)$. 假设 $\lambda_i=\lambda_i^0$ 的双边检验有 (渐近) 接受区域

$$-z(\varepsilon)\leqslant\sqrt{\frac{n}{2}}\frac{l_i-\lambda_i^0}{\lambda_i^0}\leqslant z(\varepsilon),\tag{4}$$

其中 $N(0,1)$ 分布中值大于 $z(\varepsilon)$ 的概率是 $\frac{1}{2}\varepsilon$. 区间 (4) 可以转化为 λ_i 的置信度为 $1-\varepsilon$ 的置信区间:

$$\frac{l_i}{1+\sqrt{2/n}z(\varepsilon)}\leqslant\lambda_i\leqslant\frac{l_i}{1-\sqrt{2/n}z(\varepsilon)}.\tag{5}$$

需要指出的是, 置信系数应该取得足够大, 从而 $\sqrt{2/n}z(\varepsilon) < 1$. 另外, 我们可以利用由定理 4.2.3 得到的 $\sqrt{n}(\ln l_i - \ln \lambda_i)$ 的极限分布是 $N(0,2)$ 的事实.

关于向量 $\boldsymbol{\beta}^{(i)}$ 的分量的推断可以基于把 $\boldsymbol{b}_i$ 近似看作均值为 $\boldsymbol{\beta}^{(i)}$ 且 (奇异) 协方差为 $1/n$ 乘以 (1) 的正态分布来进行.

11.6.2 特征向量的置信区域

我们利用样本特征向量的渐近分布来得到 $\boldsymbol{\Sigma}$ 的第 i 个特征向量的大样本置信区域 [Anderson(1963a)]. 协方差阵 (1) 可以记为

$$\boldsymbol{\beta}\boldsymbol{\Delta}_i^2\boldsymbol{\beta}' = \boldsymbol{\beta}_i^*\boldsymbol{\Delta}_i^{*2}\boldsymbol{\beta}_i^{*'}, \tag{6}$$

其中 $\boldsymbol{\Delta}_i$ 是 $p\times p$ 的对角矩阵, 其第 i 个对角元素为 0, 第 j 个对角元素为 $\sqrt{\lambda_i\lambda_j}/(\lambda_i-\lambda_j)$, $i\neq j$; $\boldsymbol{\Delta}_i^*$ 是从 $\boldsymbol{\Delta}$ 去掉第 i 行和第 i 列得到的 $(p-1)\times(p-1)$ 的矩阵; $\boldsymbol{\beta}_i^*$ 是从 $\boldsymbol{\beta}$ 去掉第 i 列得到的 $p\times(p-1)$ 矩阵. 则 $\boldsymbol{h}^{(i)} = \boldsymbol{\Delta}_i^{*-1}\boldsymbol{\beta}_i^{*'}\sqrt{n}\,(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)})$ 有极限正态分布, 均值为 $\boldsymbol{0}$, 协方差阵为

$$\mathrm{Cov}(\boldsymbol{h}^{(i)}) = \boldsymbol{\Delta}_i^{*-1}\boldsymbol{\beta}_i^{*'}(\boldsymbol{\beta}_i^*\boldsymbol{\Delta}_i^{*2}\boldsymbol{\beta}_i^{*'})\boldsymbol{\beta}_i^*\boldsymbol{\Delta}_i^{*-1} = \boldsymbol{I}_{p-1}, \tag{7}$$

且

$$\boldsymbol{h}^{(i)'}\boldsymbol{h}^{(i)} = n(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)})'\boldsymbol{\beta}_i^*\boldsymbol{\Delta}^{*-2}\boldsymbol{\beta}_i^{*'}(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)}) \tag{8}$$

有极限 χ^2 分布, 自由度为 $p-1$. 因为 $\boldsymbol{\beta}\boldsymbol{\Lambda}^{-1}\boldsymbol{\beta}' = \boldsymbol{\Sigma}^{-1}$,$\boldsymbol{\beta}\boldsymbol{\beta}' = \boldsymbol{I}$, $\boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}' = \boldsymbol{\Sigma}$, 所以在 $\sqrt{n}(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)})$ 中的二次型的矩阵是

$$\begin{aligned}\boldsymbol{\beta}_i^*\boldsymbol{\Delta}_i^{*-2}\boldsymbol{\beta}_i^{*'} &= \sum_{j=1}^{p}\boldsymbol{\beta}^{(j)}\left(\frac{\lambda_i}{\lambda_j}-2+\frac{\lambda_j}{\lambda_i}\right)\boldsymbol{\beta}^{(j)'} - \boldsymbol{\beta}^{(i)}\left(\frac{\lambda_i}{\lambda_i}-2+\frac{\lambda_i}{\lambda_i}\right)\boldsymbol{\beta}^{(i)'} \\ &= \lambda_i\boldsymbol{\Sigma}^{-1} - 2\boldsymbol{I} + (1/\lambda_i)\boldsymbol{\Sigma}\end{aligned} \tag{9}$$

于是因为 $\boldsymbol{\beta}^{(i)'}$ 是 $\boldsymbol{\Sigma}$ 的根 λ_i 的特征向量, 且是 $\boldsymbol{\Sigma}^{-1}$ 的根 $1/\lambda_i$ 的特征向量, (8) 式是

$$\begin{aligned}&n(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)})'\left[\lambda_i\boldsymbol{\Sigma}^{-1}-2\boldsymbol{I}+(1/\lambda_i)\boldsymbol{\Sigma}\right](\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)}) \\ =\,&n\boldsymbol{b}^{(i)'}\left[\lambda_i\boldsymbol{\Sigma}^{-1}-2\boldsymbol{I}+(1/\lambda_i)\boldsymbol{\Sigma}\right]\boldsymbol{b}^{(i)} \\ =\,&n\left[\lambda_i\boldsymbol{b}^{(i)'}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}^{(i)}+(1/\lambda_i)\boldsymbol{b}^{(i)'}\boldsymbol{\Sigma}\boldsymbol{b}^{(i)}-2\right].\end{aligned} \tag{10}$$

在 (10) 式的左边我们可以把 $\boldsymbol{\Sigma}$ 和 λ_i 替换为它们的相容估计量 $\boldsymbol{S}$ 和 l_i 来得到

$$\begin{aligned}&n(\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)})'\left[l_i\boldsymbol{S}^{-1}-2\boldsymbol{I}+(1/l_i)\boldsymbol{S}\right](\boldsymbol{b}^{(i)}-\boldsymbol{\beta}^{(i)}) \\ =\,&n\left[l_i\boldsymbol{\beta}^{(i)'}\boldsymbol{S}^{-1}\boldsymbol{\beta}^{(i)}+(1/l_i)\boldsymbol{\beta}^{(i)'}\boldsymbol{S}\boldsymbol{\beta}^{(i)}-2\right],\end{aligned} \tag{11}$$

它有极限 χ^2 分布, 自由度为 $p-1$.

$\boldsymbol{\Sigma}$ 的第 i 个特征向量的置信度为 $1-\varepsilon$ 的置信区域由 $\boldsymbol{\beta}^{(i)'}\boldsymbol{\beta}^{(i)}=1$ 和使得 (11) 式右边小于 $\chi^2_{p-1}(\varepsilon)$ 的 $\boldsymbol{\beta}^{(i)}$ 的集合的交集组成, 其中 $\Pr\left\{\chi^2_{p-1} > \chi^2_{p-1}(\varepsilon)\right\} = \varepsilon$. 需要注意的是, 二次型 (9) 的矩阵是半正定的.

这个方法也给出了第 i 个特征向量是特定的 $\beta_0^{(i)}$ ($\beta_0^{(i)\prime}\beta_0^{(i)}=1$) 的原假设的一个检验. 如果 (11) 式的右边用 $\beta_0^{(i)}$ 代替 $\beta^{(i)}$ 后超出了 $\chi^2_{p-1}(\varepsilon)$ 则拒绝假设.

Mallows(1961) 给出了 $\boldsymbol{\Sigma}$ 的某个特征向量是否为 β_0 的检验. 令 $\mathbf{B}_0$ 是 $p\times(p-1)$ 的矩阵满足 $\beta_0'\mathbf{B}_0=\mathbf{0}$. 如果原假设为真, 则 $\beta_0'\boldsymbol{X}$ 和 $\mathbf{B}_0'\boldsymbol{X}$ 是独立的 (因为 $\mathbf{B}_0$ 是其他特征向量集的非奇异变换). 这个检验是基于 $\beta_0'\boldsymbol{X}$ 和 $\mathbf{B}_0'\boldsymbol{X}$ 的多重相关的. 本质上, 这个检验方法可以用来得到置信区域. 这些方法的使用受到假设向量依附特征根的限制; 解释依赖于根 (例如, 最大的和最小的).

Tyler(1981, 1983b) 推广了置信区域 (11), 使得它包含线性子空间中的向量. 他也研究了关于解除总体服从正态分布的限制问题.

11.6.3 特征根的精确置信限

现在我们考虑 $\boldsymbol{\Sigma}$ 的所有特征根 $\lambda_1\geqslant\cdots\geqslant\lambda_p$ 的一个置信区间 [Anderson (1965a)]. 我们利用事实 $\boldsymbol{\beta}^{(i)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(i)}=\lambda_i,\boldsymbol{\beta}^{(i)\prime}\boldsymbol{\beta}^{(i)}=1,i=1,p$, 和 $\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(p)}=0=\boldsymbol{\beta}^{(1)\prime}\boldsymbol{\beta}^{(p)}$. 则 $\boldsymbol{\beta}^{(1)\prime}\boldsymbol{X}$ 和 $\boldsymbol{\beta}^{(p)\prime}\boldsymbol{X}$ 是不相关的且分别有方差 λ_1 和 λ_p. 因此 $n\boldsymbol{\beta}^{(1)\prime}\boldsymbol{S}\boldsymbol{\beta}^{(1)}/\lambda_1$ 和 $n\boldsymbol{\beta}^{(p)\prime}\boldsymbol{S}\boldsymbol{\beta}^{(p)}/\lambda_p$ 是独立地服从自由度为 n 的 χ^2 分布的. 令 l 和 u 是满足

$$1-\varepsilon=\Pr\left\{nl\leqslant\chi^2_n\right\}\Pr\left\{\chi^2_n\leqslant nu\right\}\tag{12}$$

的两个数. 则

$$\begin{aligned}1-\varepsilon&=\Pr\left\{l\leqslant\frac{\boldsymbol{\beta}^{(1)\prime}\boldsymbol{S}\boldsymbol{\beta}^{(1)}}{\lambda_1},\frac{\boldsymbol{\beta}^{(p)\prime}\boldsymbol{\Sigma}\boldsymbol{\beta}^{(p)}}{\lambda_p}\leqslant u\right\}\\&=\Pr\left\{\frac{\boldsymbol{\beta}^{(p)\prime}\boldsymbol{S}\boldsymbol{\beta}^{(p)}}{u}\leqslant\lambda_p,\lambda_1\leqslant\frac{\boldsymbol{\beta}^{(1)\prime}\boldsymbol{S}\boldsymbol{\beta}^{(1)}}{l}\right\}\\&\leqslant\Pr\left\{\min_{\boldsymbol{b}'\boldsymbol{b}=1}\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{u}\leqslant\lambda_p,\lambda_1\leqslant\max_{\boldsymbol{b}'\boldsymbol{b}=1}\frac{\boldsymbol{b}'\boldsymbol{S}\boldsymbol{b}}{l}\right\}\\&=\Pr\left\{\frac{l_p}{u}\leqslant\lambda_p\leqslant\lambda_1\leqslant\frac{l_1}{l}\right\}.\end{aligned}\tag{13}$$

定理 11.6.1 *$\boldsymbol{\Sigma}$ 的特征根的置信度至少为 $1-\varepsilon$ 的置信区间是*

$$l_p/u\leqslant\lambda_p\leqslant\lambda_1\leqslant l_1/l,\tag{14}$$

其中 l 和 u 满足 (12) 式.

更紧凑的不等式可以得出更好的下界. 因为 $\mathbf{B}$ 是正交的, 所以矩阵 $\boldsymbol{H}=n\mathbf{B}'\boldsymbol{S}\mathbf{B}$ 有特征根 $nl_1,\cdots,nl_p$. 我们利用下面的引理.

引理 11.6.1 *对于任意一个正定矩阵 $\boldsymbol{H}$*

$$\mathrm{ch}_p(\boldsymbol{H})\leqslant\frac{1}{h^{ii}}\leqslant\mathrm{ch}_1(\boldsymbol{H}),\quad i=1,\cdots,p,\tag{15}$$

其中 $\boldsymbol{H}^{-1}=(h^{ij})$, $\mathrm{ch}_p(\boldsymbol{H})$ 和 $\mathrm{ch}_1(\boldsymbol{H})$ 分别是 $\boldsymbol{H}$ 的最小特征根和最大特征根.

证明 由附录中的定理 A.2.4, 我们知道 $\mathrm{ch}_p(\boldsymbol{H})\leqslant h_{ii}\leqslant \mathrm{ch}_1(\boldsymbol{H})$ 且

$$\mathrm{ch}_p(\boldsymbol{H}^{-1})\leqslant h^{ii}\leqslant \mathrm{ch}_1(\boldsymbol{H}^{-1}),\quad i=1,\cdots,p. \tag{16}$$

因为 $\mathrm{ch}_p(\boldsymbol{H})=1/\mathrm{ch}_1(\boldsymbol{H}^{-1})$ 且 $\mathrm{ch}_1(\boldsymbol{H})=1/\mathrm{ch}_p(\boldsymbol{H}^{-1})$, 所以这个引理成立. ■

定理 5.2.2 的论述表明 $1/(\lambda_p h^{pp})$ 服从自由度为 $n-p+1$ 的 χ^2 分布, 而定理 4.3.3 表明 h^{pp} 和 h_{11} 是独立的. 令 l' 和 u' 是满足

$$1-\varepsilon=\Pr\left\{nl'\leqslant\chi_n^2\right\}\Pr\left\{\chi_{n-p+1}^2\leqslant nu'\right\} \tag{17}$$

的两个数. 则由于 $\mathrm{ch}_p(\boldsymbol{H})=nl_p$ 且 $\mathrm{ch}_1(\boldsymbol{H})=nl_1$,

$$\begin{aligned}1-\varepsilon&=\Pr\left\{nl'\leqslant\frac{h_{11}}{\lambda_1},\frac{1}{\lambda_p h^{pp}}\leqslant nu'\right\} \qquad (18)\\ &\leqslant\Pr\left\{\frac{l_p}{u'}\leqslant\lambda_p,\lambda_1\leqslant\frac{l_1}{l'}\right\}.\end{aligned}$$

定理 11.6.2 $\boldsymbol{\Sigma}$ 的特征根的置信度至少为 $1-\varepsilon$ 的置信区间是

$$\frac{l_p}{u'}\leqslant\lambda_p\leqslant\lambda_1\leqslant\frac{l_1}{l'}, \tag{19}$$

其中 l' 和 u' 满足 (17) 式.

Anderson(1965a, 1965b) 指出上面的置信界在界类

$$f(l_1,\cdots,l_p)\leqslant\lambda_p\leqslant\lambda_1\leqslant g(l_1,\cdots,l_p) \tag{20}$$

中是最优的, 其中 f 和 g 是 1 次齐性的, 并且在其他自变量不变时 f 和 g 关于每个自变量是单调非减的. 如果 (20) 至少以概率 $1-\varepsilon$ 成立, 则存在一对数 u' 和 l' 满足 (17) 式且

$$f(l_1,\cdots,l_p)\leqslant\frac{l_p}{u'},\quad \frac{l_1}{l'}\leqslant g(l_1,\cdots,l_p). \tag{21}$$

齐性条件意味着, 如果观测向量乘以 c, 则置信界乘以 c^2(这是一种尺度不变性). 单调性条件表示增加 $\boldsymbol{S}$ 的大小会导致 $\boldsymbol{\Sigma}$ 的极限的增加 (这是一种相容性).

10.8 节中 (31) 式给出的 $\boldsymbol{\Sigma}$ 的根的置信界更好, 它是基于 $\boldsymbol{\Sigma}=\boldsymbol{I}$ 时 $\boldsymbol{S}$ 的根的分布的.

11.7 关于协方差阵的特征根的假设检验

11.7.1 检验关于最小的特征根的和的假设

研究者可能会问最后 $p-m$ 个主成分是否可以忽略, 即前面的 m 个主成分是否可以提供 $\boldsymbol{X}$ 的一个好的近似. 如果最后几个主成分的方差的和小于一个特定的量 (比如 γ) 的话, 那么他就可以这样做. 考虑原假设

$$H:\lambda_{m+1}+\cdots+\lambda_p\geqslant\gamma, \tag{1}$$

其中 γ 是确定的, 相应地备择假设是和小于 γ. 如果 $\boldsymbol{\Sigma}$ 的特征根是不同的, 则由定理 13.5.1 得

$$\sqrt{n}\left(\sum_{i=m+1}^{p}l_i-\sum_{i=m+1}^{p}\lambda_i\right) \tag{2}$$

有极限正态分布, 均值为 0, 方差为 $2\sum_{i=m+1}^{p}\lambda_i^2$. 方差可以用 $2\sum_{i=m+1}^{p}l_i^2$ 来相容地估计. 则 (大样本) 显著性水平为 ε 的拒绝区域是

$$\sum_{i=m+1}^{p}l_i<\gamma-\frac{\sqrt{2\sum_{i=m+1}^{p}l_i^2}}{\sqrt{n}}z(2\varepsilon), \tag{3}$$

其中 $z(2\varepsilon)$ 是标准正态分布的显著性水平为 ε 的上分位数. 如果 (1) 中等式成立, 则 (大样本) 拒绝概率是 ε, 如果不等式成立, 则 (大样本) 拒绝概率小于 ε.

研究者可能还想要 $\sum_{i=m+1}^{p}\lambda_i$ 的近似置信水平至少为 $1-\varepsilon$ 的上置信区间. 它是

$$\sum_{i=m+1}^{p}\lambda_i\leqslant\sum_{i=m+1}^{p}l_i+\frac{\sqrt{2\sum_{i=m+1}^{p}l_i^2}}{\sqrt{n}}z(2\varepsilon). \tag{4}$$

如果右边足够小 (特别是小于 γ 时), 则研究者有信心说最小的 $p-m$ 个主成分的方差和太小了, 以至于可以忽略它们. Anderson(1963a) 也给出了当 $\lambda_{m+1}=\cdots=\lambda_p$ 时的这种分析.

11.7.2 检验关于最小的特征根的和相对于所有根的和的假设

如果最后 $p-m$ 个主成分的和相对于所有根的和 (也就是协方差阵的迹) 太小, 则研究者可能会忽略它们. 考虑原假设

$$H:f(\boldsymbol{\lambda})=\frac{\lambda_{m+1}+\cdots+\lambda_p}{\lambda_1+\cdots+\lambda_p}\geqslant\delta, \tag{5}$$

其中 δ 是确定的, 相应的备择假设是 $f(\boldsymbol{\lambda})<\delta$. 我们利用下面的事实,

$$\begin{aligned}\frac{\partial f(\lambda)}{\partial\lambda_i}&=-\frac{\lambda_{m+1}+\cdots+\lambda_p}{(\lambda_1+\cdots+\lambda_p)^2},\quad i=1,\cdots,m,\\ \frac{\partial f(\lambda)}{\partial\lambda_i}&=\frac{\lambda_1+\cdots+\lambda_m}{(\lambda_1+\cdots+\lambda_p)^2},\quad i=m+1,\cdots,p.\end{aligned} \tag{6}$$

则当 (5) 式中等式成立, 由定理 4.2.3 知 $f(\boldsymbol{l})$ 的渐近方差是

$$2\left(\frac{\delta}{\operatorname{tr}\boldsymbol{\Sigma}}\right)^2(\lambda_1^2+\cdots+\lambda_m^2)+2\left(\frac{1-\delta}{\operatorname{tr}\boldsymbol{\Sigma}}\right)^2(\lambda_{m+1}^2+\cdots+\lambda_p^2). \tag{7}$$

如果 $\sqrt{n}[f(\boldsymbol{l})-\delta]$ 小于标准正态分布的合适分位数乘以 (7) 式的根, 其中所有的 λ 都用 l 代替, 而 $\operatorname{tr}\boldsymbol{\Sigma}$ 用 $\operatorname{tr}\boldsymbol{S}$ 代替, 则拒绝原假设 H. 另外我们可以构造 $f(\boldsymbol{\lambda})$ 的一个大样本置信区域. 近似置信水平为 $1-\varepsilon$ 的置信区域是 $[z=z(2\varepsilon)]$

$$\frac{\sum_{i=m+1}^{p}\lambda_i}{\sum_{i=1}^{p}\lambda_i}\leqslant\frac{\sum_{i=m+1}^{p}l_i}{\sum_{i=1}^{p}l_i}+z\frac{\left[2(\sum_{i=m+1}^{p}l_i)^2\sum_{i=1}^{m}l_i^2+2(\sum_{i=1}^{m}l_i)^2\sum_{i=m+1}^{p}l_i^2\right]^{\frac{1}{2}}}{\sqrt{n}(\sum_{i=1}^{p}l_i)^2}. \tag{8}$$

如果右边足够小, 研究者就会用前面几个主成分来代替所有的观测向量.

11.7.3 最小根的等价检验

假设观测 $\boldsymbol{X}$ 由 $\boldsymbol{V}+\boldsymbol{U}+\boldsymbol{\mu}$ 给出, 其中 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 是均值为 0 的不可观测的随机向量, $\boldsymbol{\mu}$ 是不可观测的固定向量. 如果 $\mathrm{E}(\boldsymbol{U}\boldsymbol{U}')=\sigma^2\boldsymbol{I}$, 则 $\boldsymbol{U}$ 可以解释为由不相关且等方差的观测误差组成的. (假设 $\boldsymbol{X}$ 的所有分量用相同的度量单位.) 而 $\boldsymbol{V}$ 可以解释为由系统部分组成, 且是在一个 m 维空间中. 则 $\mathrm{E}(\boldsymbol{V}\boldsymbol{V}')=\boldsymbol{\Phi}$ 是秩为 m 的半正定矩阵. 观测协方差阵 $\boldsymbol{\Sigma}=\boldsymbol{\Phi}+\sigma^2\boldsymbol{I}$ 有一个 $p-m$ 重的特征根 σ^2(习题 11.4).

在这一小节中我们考虑检验原假设 $\lambda_{m+1}=\cdots=\lambda_p$. 这等价于原假设 $\boldsymbol{\Sigma}=\boldsymbol{\Phi}+\sigma^2\boldsymbol{I}$, 其中 $\boldsymbol{\Phi}$ 是秩为 m 且半正定的. 在 10.7 节中, 我们看到当 $m=0$ 时, 似然比准则是样本根的几何平均值和算术平均值的比的 $\frac{1}{2}pN$ 次幂. 这里类似的准则是

$$\frac{\prod_{i=m+1}^{p}l_i}{(\sum_{i=m+1}^{p}l_i)^{p-m}}(p-m)^{p-m} \tag{9}$$

的 $\frac{1}{2}N$ 次幂. 这也是似然比准则, 但是我们不推导它. [参见 Anderson(1963a).] 令 $\sqrt{n}(l_i-\lambda_{m+1})=d_i$, $i=m+1,\cdots,p$. 在原假设下 (9) 式的对数乘以 $-n$ 和下式近似相等

$$\begin{aligned}
&-n\ln\prod_{i=m+1}^{p}l_i+n(p-m)\ln\frac{\sum_{i=m+1}^{p}l_i}{p-m} \qquad (10)\\
=&-n\sum_{i=m+1}^{p}\ln(\lambda_{m+1}+n^{-\frac{1}{2}}d_i)+n(p-m)\ln\frac{\sum_{i=m+1}^{p}(\lambda_{m+1}+n^{-\frac{1}{2}}d_i)}{p-m}\\
=&n\left\{-\sum_{i=m+1}^{p}\ln\left(1+\frac{d_i}{\lambda_{m+1}n^{\frac{1}{2}}}\right)+(p-m)\ln\left(1+\frac{\sum_{i=m+1}^{p}d_i}{(p-m)\lambda_{m+1}n^{\frac{1}{2}}}\right)\right\}\\
=&n\left\{-\sum_{i=m+1}^{p}\left[\frac{d_i}{\lambda_{m+1}n^{\frac{1}{2}}}-\frac{d_i^2}{2\lambda_{m+1}^2n}+\cdots\right]\right.\\
&\left.+(p-m)\left[\frac{\sum_{i=m+1}^{p}d_i}{(p-m)\lambda_{m+1}n^{\frac{1}{2}}}-\frac{(\sum_{i=m+1}^{p}d_i)^2}{2(p-m)^2\lambda_{m+1}^2n}+\cdots\right]\right\}\\
=&n\left\{\sum_{i=m+1}^{p}\frac{d_i^2}{2\lambda_{m+1}^2n}-\cdots-\frac{(\sum_{i=m+1}^{p}d_i)^2}{2(p-m)\lambda_{m+1}^2n}+\cdots\right\}\\
=&\frac{1}{2\lambda_{m+1}^2}\left[\sum_{i=m+1}^{p}d_i^2-\frac{1}{p-m}\left(\sum_{i=m+1}^{p}d_i\right)^2\right]+O_p(1).
\end{aligned}$$

在 13.5.2 节中将证明 $d_{m+1},\cdots,d_p$ 的极限分布和对称矩阵 $\boldsymbol{U}_{22}=(u_{ij})\,(i,j=m+1,\cdots,p)$ 的根的分布相同, U_{22} 的函数独立元素是独立的且服从均值为 0 的正态分布; 其非对角元素 $u_{ij}\,(i<j)$ 有方差 λ_{m+1}^2, 对角元素 u_{ii} 有方差 $2\lambda_{m+1}^2$. 见定理 13.5.2. 则 (10) 的极限分布是

$$\begin{aligned}&\frac{1}{2\lambda_{m+1}^2}\left(\operatorname{tr}\boldsymbol{U}_{22}^2-\frac{1}{p-m}(\operatorname{tr}\boldsymbol{U}_{22})^2\right)\\&=\frac{1}{2\lambda_{m+1}^2}\left[\operatorname{tr}\boldsymbol{U}_{22}\boldsymbol{U}_{22}'-\frac{1}{p-m}(\operatorname{tr}\boldsymbol{U}_{22})^2\right]\\&=\frac{1}{2\lambda_{m+1}^2}\left[2\sum_{i<j}u_{ij}^2+\sum_{i=m+1}^{p}u_{ii}^2-\frac{1}{p-m}\left(\sum_{i=m+1}^{p}u_{ii}\right)^2\right].\end{aligned}\tag{11}$$

因此 $\sum_{i<j}u_{ij}^2/\lambda_{m+1}^2$ 是渐近地服从自由度为 $\frac{1}{2}(p-m)(p-m-1)$ 的 χ^2 分布的; $\frac{1}{2}[\sum_{i=m+1}^{p}u_{ii}^2-(\sum_{i=m+1}^{p}u_{ii})^2/(p-m)]/\lambda_{m+1}^2$ 是渐近地服从自由度为 $p-m-1$ 的 χ^2 分布的. 因此 (10) 有自由度为 $\frac{1}{2}(p-m+2)(p-m-1)$ 的极限 χ^2 分布. 如果 (10) 式的左边大于 χ^2 分布的上分位数, 则拒绝假设. 如果假设没有被拒绝, 则研究者可以认为最后 $p-m$ 个主成分是完全由误差组成的.

当观测的单位不完全相同时, 11.7 节中考虑的三个假设的意义就有问题. 相应的对于相关矩阵的假设的解释也就令人怀疑. 而且, 最后的准则 (通常) 不再有 χ^2 分布. 更多的讨论由 Anderson(1963a) 给出.

准则 (9) 对应于 10.7 节中的球形准则, 并且相应的 χ^2 分布的自由度为 $\frac{1}{2}(p-m)(p-m+1)-1$.

11.8 椭球等高分布

11.8.1 椭球等高观测

令 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是密度为

$$|\boldsymbol{\Psi}|^{-\frac{1}{2}}g\left[(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})\right]\tag{1}$$

的随机向量 $\boldsymbol{X}$ 的 N 个观测, 其中 $\boldsymbol{\Psi}$ 是一个正定矩阵, $R^2=(\boldsymbol{x}-\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{\nu})$, 且 $\mathrm{E}(R^4)<\infty$. 定义 $\kappa=p\mathrm{E}(R^4)/[(\mathrm{E}(R^2))^2(p+2)]-1$. 则 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{\nu}=\boldsymbol{\mu}$ 且 $\mathrm{E}(\boldsymbol{X}-\boldsymbol{\nu})(\boldsymbol{X}-\boldsymbol{\nu})'=(\mathrm{E}(R^2)/p)\boldsymbol{\Psi}=\boldsymbol{\Sigma}$.

$\boldsymbol{\Sigma}$ 的主成分的极大似然估计是由 3.6 节中 (20) 式给出的 $\hat{\boldsymbol{\Sigma}}=(\mathrm{E}(R^2)/p)\hat{\boldsymbol{\Lambda}}$ 的特征根和特征向量. 另外还可以用 $\boldsymbol{S}$ 的特征根和特征向量来估计, 其中 $\boldsymbol{S}$ 是 $\boldsymbol{\Sigma}$ 的无偏估计. 这些估计的渐近正态分布将在 13.7 节中得到 (定理 13.7.1). 令 $\boldsymbol{\Sigma}=\boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}'$ 且 $\boldsymbol{S}=\boldsymbol{B}\boldsymbol{L}\boldsymbol{B}'$, 其中 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{L}$ 是对角的, 而 $\boldsymbol{\beta}$ 和 $\boldsymbol{B}$ 是正交的. 令 $\boldsymbol{D}=\sqrt{N}(\boldsymbol{L}-\boldsymbol{\Lambda})$ 且 $\boldsymbol{G}=\sqrt{N}(\boldsymbol{B}-\boldsymbol{\beta})$. 则 $\boldsymbol{D}$ 和 $\boldsymbol{G}$ 的极限分布是正态分布且 $\boldsymbol{G}$ 和 $\boldsymbol{D}$ 是独立的.

d_i 的方差是 $(2+3\kappa)\lambda_i^2$, 且 d_i 和 $d_j\ (i \neq j)$ 的协方差是 $\kappa\lambda_i\lambda_j$. $\boldsymbol{g}_i$ 的协方差是

$$\mathrm{Cov}(\boldsymbol{g}_i) = (1+\kappa)\sum_{\substack{k=1\\k\neq i}}^{p} \frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\boldsymbol{\beta}_k\boldsymbol{\beta}_k'. \tag{2}$$

$\boldsymbol{g}_i$ 和 $\boldsymbol{g}_j$ 的协方差是

$$\mathrm{Cov}(\boldsymbol{g}_i, \boldsymbol{g}_j) = -(1+\kappa)\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_i'. \tag{3}$$

关于单个有序根 λ_i 的推断可以利用 $\sqrt{N}(l_i-\lambda_i)/(\sqrt{2(2+3\hat{\kappa})}l_i)$ 的极限分布是标准正态分布.

关于单个向量的推断可以利用 11.6.2 节中 (11) 式的右边, 其中 $\boldsymbol{S}$ 用 $(1+\hat{\kappa})\boldsymbol{S}$ 来代替, $\boldsymbol{S}^{-1}$ 用 $\boldsymbol{S}^{-1}/(1+\hat{\kappa})$ 来代替.

在 13.7.1 节中将证明, 检验 $q=p-m$ 个最小根相等的似然比准则的对数的极限分布是 $(1+\kappa)\chi^2_{q(q-1)/2-1}$ 的分布.

11.8.2 椭球等高矩阵分布

假设 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 的密度是

$$\begin{aligned}&|\boldsymbol{\Psi}|^{-N/2}g\left[\mathrm{tr}\,(\boldsymbol{X}-\boldsymbol{\varepsilon}_N'\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\boldsymbol{X}-\boldsymbol{\varepsilon}_N'\boldsymbol{\nu})\right]\\ =&|\boldsymbol{\Psi}|^{-N/2}g\left[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Psi}^{-1}+N(\bar{\boldsymbol{x}}-\boldsymbol{\nu})'\boldsymbol{\Psi}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\nu})\right],\end{aligned}$$

其中 $\boldsymbol{A}=(\boldsymbol{X}-\boldsymbol{\varepsilon}_N'\bar{\boldsymbol{x}})(\boldsymbol{X}-\boldsymbol{\varepsilon}_N'\bar{\boldsymbol{x}})'=n\boldsymbol{S}$ 且 $n=N-1$. 因此 $\boldsymbol{x}$ 和 $\boldsymbol{A}$ 是统计量的一个充分集.

现在考虑 $\boldsymbol{A}=\boldsymbol{Y}\boldsymbol{Y}'$ 有密度 $g(\mathrm{tr}\ \boldsymbol{A})$ 的情形. 令 $\boldsymbol{A}=\boldsymbol{B}\boldsymbol{L}\boldsymbol{B}'$, 其中 $\boldsymbol{L}$ 是对角的且对角元素是 $l_1>\cdots>l_p$, 而 $\boldsymbol{B}$ 是正交的且 $p_{i1}\geqslant 0$. 则 $\boldsymbol{L}$ 和 $\boldsymbol{B}$ 是独立的; 根 $l_1,\cdots,l_p$ 有 13.7 节中的密度 (18), 且矩阵 $\boldsymbol{B}$ 有条件 Haar 不变分布.

习 题

11.1 (11.2 节) 证明 $\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$ 的特征向量是

$$\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix} \quad 和 \quad \begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix},$$

分别对应于根 $1+\rho$ 和 $1-\rho$.

11.2 (11.2 节) 对于任意一个实对称矩阵, 验证由定理 11.2.1 的证明可以得出附录中定理 A.2.1 的证明.

11.3 (11.2 节) 令 $\boldsymbol{z}=\boldsymbol{y}+\boldsymbol{x}$, 其中 $\mathrm{E}(\boldsymbol{y})=\mathrm{E}(\boldsymbol{x})=\boldsymbol{0}, \mathrm{E}(\boldsymbol{y}\boldsymbol{y}')=\boldsymbol{\Phi}$, $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')=\sigma^2\boldsymbol{I}$, $\mathrm{E}(\boldsymbol{y}\boldsymbol{x}')=\boldsymbol{0}$. $\boldsymbol{y}$ 的 p 个分量称为系统部分, $\boldsymbol{x}$ 的分量称为误差.

(a) 求有单位方差的线性组合 $\gamma' z$ 使得它有极小的误差方差 (即 $\gamma' x$ 有极小方差).

(b) 假设 $\phi_{ii} + \sigma^2 = 1, i = 1, \cdots, p$. 求具有单位方差的线性函数 $\gamma' z$, 使得 z_i 和 $\gamma' z$, $i = 1, \cdots, p$ 的相关系数的平方和极大化.

(c) 把这些结果和主成分相联系.

11.4 (11.2 节) 令 $\Sigma = \Phi + \sigma^2 I$, 其中 Φ 是秩为 m 的半正定矩阵. 证明 Φ 的每个特征向量也是 Σ 的特征向量, 且 Φ 的根也都是 Σ 的根加上 σ^2.

11.5 (11.2 节) 令 Σ 的特征根是 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p \geqslant 0$.

(a) 如果 $\lambda_1 = \lambda_2 = \cdots = \lambda_p > 0$, 则 Σ 的形式是怎样的? 一个常密度的椭球面是什么形状的?

(b) 如果 $\lambda_1 > \lambda_2 = \cdots = \lambda_p > 0$, 则 Σ 的形式是怎样的? 一个常密度的椭球面是什么形状的?

(c) 如果 $\lambda_1 = \lambda_2 = \cdots = \lambda_{p-1} > \lambda_p > 0$, 则 Σ 的形式是怎样的? 一个常密度的椭球面是什么形状的?

11.6 (11.2 节) *组内相关*. 令

$$\Sigma = \sigma^2[(1-\rho)I + \rho\varepsilon\varepsilon'],$$

其中 $\varepsilon = (1, \cdots, 1)'$. 证明对于 $\rho > 0$, 最大的特征根是 $\sigma^2[1 + (p-1)\rho]$ 且相应的特征向量是 ε. 证明如果 $\varepsilon' x = 0$, 则 x 是对应于根 $\sigma^2(1-\rho)$ 的特征向量. 证明根 $\sigma^2(1-\rho)$ 的重数为 $p-1$.

11.7 (11.3 节) 在 9.7 节的例子中, 考虑三个压制操作 (x_2, x_4, x_5). 求这个估计协方差阵的第一个主成分. [提示: 由向量 (1, 1, 1) 开始并迭代.]

11.8 (11.3 节) 直接证明定理 11.2.1 的样本形式, 其中 $\sum x_\alpha = \mathbf{0}$, $\sum x_\alpha x_\alpha' = A$.

11.9 (11.3 节) 令 l_1 和 l_p 分别是 S 的最大特征根和最小特征根. 证明 $\mathrm{E}(l_1) \geqslant \lambda_1$ 且 $\mathrm{E}(l_p) \leqslant \lambda_p$.

11.10 (11.3 节) 令 $U_1 = \beta^{(1)\prime} X$ 是有方差 $\mathrm{Var}(U_1) = \lambda_1$ 的第一个总体主成分, 且令 $V_1 = b^{(1)\prime} X$ 是有 (样本) 方差 (基于 S) l_1 的第一个样本主成分. 令 S^* 是另一个 (独立的) 样本的协方差阵. 证明 $\mathrm{E} b^{(1)\prime} S^* b^{(1)} \leqslant \lambda_1$.

11.11 (11.3 节) 假设对于每一个 $i, j, \sigma_{ij} > 0\,[\Sigma = (\sigma_{ij})]$. 证明

(a) 第一个主成分的系数的符号相同,

(b) 其余主成分的系数不能同时同号.

11.12 (11.4 节) 当 $\lambda_1 > \lambda_2$ 时证明 (4).

(a) 证明 $\Sigma^i = \boldsymbol{\beta}\Lambda^i\boldsymbol{\beta}'$.

(b) 证明

$$y_{(i)} = t_i \boldsymbol{\beta}\Lambda^i\boldsymbol{\beta}' x_{(0)} = t_i \lambda_1^i \boldsymbol{\beta} \left(\frac{1}{\lambda_1}\Lambda\right)^i \boldsymbol{\beta}' x_{(0)},$$

其中 $t_i = \prod_{j=0}^{i} s_j$ 且 $s_j = 1/\sqrt{x_{(i)}' x_{(i)}}$.

(c) 证明

$$\lim_{i\to\infty} \left(\frac{1}{\lambda_1}\Lambda\right)^i = E_{11},$$

其中 E_{11} 在左上角的位置为 1 而其他位置为 0.

(d) 证明 $\lim_{i\to\infty}(t_i\lambda_1^i)^2 = 1/(\beta^{(1)\prime}x_{(0)})^2$.

(e) 总结这些证明.

11.13 (11.4 节) 令

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma'_{(1)} \\ \sigma_{(1)} & \Sigma_{22} \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 \\ 0 & H \end{bmatrix},$$

其中 $H = I_{p-1} - 2\alpha\alpha'$ 且 α 有 $p-1$ 个分量. 证明α可以这样选择, 使得在

$$K\Sigma K = \begin{bmatrix} \sigma_{11} & \sigma'_{(1)}H \\ H\sigma_{(1)} & H\Sigma_{22}H \end{bmatrix}$$

中 $H\sigma_{(1)}$ 除了第一个分量外其余分量都为 0.

11.14 (11.6 节) 证明

$$\ln l_i - \sqrt{\frac{2}{n}}z(\varepsilon) < \ln\lambda_i < \ln l_i + \sqrt{\frac{2}{n}}z(\varepsilon)$$

是 $\ln\lambda_i$ 的一个置信区间, 置信水平大约为 $1-\varepsilon$.

11.15 (11.6 节) 证明: 如果 $l' = l$ 且 $p > 2$ 则 $u' < u$.

11.16 (11.6 节) 证明: 如果 $l = l^*$ 且 $p > 2$ 则 $u < u^*$, 其中 l^* 和 u^* 是 10.8.4 节中的 l 和 u.

11.17 下面给出的是 24 只雄性龟的长度, 宽度和高度 (以毫米为单位)[Jolicoeur and Mosimann(1960)]. 求出 (样本) 主成分和它们的方差.

样本号	长度	宽度	高度	样本号	长度	宽度	高度
1	93	74	37	13	116	90	43
2	94	78	35	14	117	90	41
3	96	80	35	15	117	91	41
4	101	84	39	16	119	93	41
5	102	85	38	17	120	89	40
6	103	81	37	18	120	93	44
7	104	83	39	19	121	95	42
8	106	83	39	20	125	93	45
9	107	82	38	21	127	96	45
10	112	89	40	22	128	95	46
11	113	88	40	23	131	95	46
12	114	86	40	24	135	106	47

第 12 章 典型相关和典型变量

12.1 引 言

本章讨论有联合分布的两组变量，我们要分析其中一组变量和另外一组变量之间的关系. 我们在每组变量张成的空间中找到了一个新的坐标系，使得新的坐标可以更清晰地反应变量间的相关性. 更准确地说，我们找到了每组变量的线性组合，使得它们之间有极大相关；这些线性组合是新系下的第一组坐标. 然后对每组变量寻找第二组线性组合，使得它们之间的相关在这样的线性组合中达到极大，并且和我们得到的第一组线性组合是不相关的. 继续这样的步骤直到两个新的坐标系完全确定.

上面简述的统计方法尤其在探索性研究中有特殊的用处. 研究者可能有两大组变量，并且想研究它们之间的相互关系. 假如这两个组很大，研究者可能就只想考虑每组的几个线性组合. 然后研究这些线性组合中高度相关的. 举例来说，一组变量是生理特征的观测值，例如不同的人头长和头宽的数据；另外一组变量是智力特征的度量值，比如智力测试的得分. 如果研究者的兴趣在于联系二者，他会发现二者的相互关系通过前几个典型变量的相关就几乎被完全刻画了.

基本理论的发展参考 Hotelling (1935), (1936).

12.2 节定义了总体的典型相关和典型变量，它们包含了一种到典范型的线性变换. 极大似然估计量是样本的模拟. 基于渐近理论的独立性检验和相关阵的秩检验将在 12.4 节中介绍.

典型相关和典型变量的另一种形式是，一组变量是随机的，另一组变量由非随机变量组成；随机变量的期望值是非随机变量的线性组合 (12.6 节). 这是 8.2 节中的模型. 一组典型变量由随机变量的线性组合组成，另一组由非随机变量的线性组合组成；第一组变量中的元对第二组变量中的元的回归效应被极大化. 在这个结构中研究线性函数的关系.

12.7 节中介绍了联立方程模型. 这个模型中单个方程的估计在形式上等同于单个线性函数关系的估计. 还简单介绍了极限信息极大似然估计量和二级最小二乘估计量.

12.2 总体的典型相关和典型变量

假设随机向量 $\boldsymbol{X}$ 有 p 个分量, 协方差阵是 $\boldsymbol{\Sigma}$ (假设为正定). 因为本章中我们只对方差和协方差感兴趣, 所以对总体我们假设 $\mathrm{E}(\boldsymbol{X})=0$. 为了拓展概念和代数知识, 我们不需要假定 $\boldsymbol{X}$ 服从正态分布, 虽然这个假设常被用于抽样理论中.

我们把 $\boldsymbol{X}$ 分成两个子向量, 分别为 p_1 元和 p_2 元, 记为

$$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{pmatrix}. \tag{1}$$

为了方便, 我们可以假设 $p_1\leqslant p_2$. 协方差阵也相应地分成 p_1 行和 p_2 行以及 p_1 列和 p_2 列,

$$\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}. \tag{2}$$

在前面的章节中, 我们介绍了一种使得方差性质最大化体现的坐标轴旋转. 这里我们介绍一种前 p_1 个坐标轴和后 p_2 个坐标轴到新的 (p_1+p_2) 系的变换, 它将更清晰地体现 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的相互关系.

考虑 $\boldsymbol{X}^{(1)}$ 的分量的任意一个线性组合, $U=\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$ 的分量的任意一个线性函数, $V=\boldsymbol{\gamma}'\boldsymbol{X}^{(2)}$. 我们首先要让两组线性函数之间有极大相关. 由于 U 和 V 的相关同 U 的倍数和 V 的倍数之间的相关相同, 我们可以对 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 进行任意正规化. 我们需要 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 使得 U 和 V 有单位方差, 即

$$1=\mathrm{E}(U^2)=\mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}\boldsymbol{X}^{(1)\prime}\boldsymbol{\alpha})=\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}, \tag{3}$$

$$1=\mathrm{E}(V^2)=\mathrm{E}(\boldsymbol{\gamma}'\boldsymbol{X}^{(2)}\boldsymbol{X}^{(2)\prime}\boldsymbol{\gamma})=\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}. \tag{4}$$

我们注意到 $\mathrm{E}(U)=\mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X}^{(1)})=\boldsymbol{\alpha}'\mathrm{E}(\boldsymbol{X}^{(1)})=0$, 类似地, $\mathrm{E}(V)=0$. 则 U 和 V 之间的相关是

$$\mathrm{E}(UV)=\mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}\boldsymbol{X}^{(2)\prime}\boldsymbol{\gamma})=\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}. \tag{5}$$

因此代数问题是在满足 (3) 和 (4) 的条件下, 寻求使 (5) 极大化的 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$.

令

$$\psi=\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}-\frac{1}{2}\lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}-1)-\frac{1}{2}\mu(\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}-1), \tag{6}$$

其中 λ 和 μ 是拉格朗日乘数. 我们对 ψ 求关于 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 的偏导数. 令求偏导后的式子等于零, 即

$$\frac{\partial\psi}{\partial\boldsymbol{\alpha}}=\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}-\lambda\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}=\mathbf{0}, \tag{7}$$

$$\frac{\partial\psi}{\partial\boldsymbol{\gamma}}=\boldsymbol{\Sigma}_{12}'\boldsymbol{\alpha}-\mu\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}=\mathbf{0}. \tag{8}$$

将 (7) 式左端乘 $\boldsymbol{\alpha}'$, (8) 式左端乘 $\boldsymbol{\gamma}'$, 得到

$$\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}-\lambda\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}=0, \tag{9}$$

$$\gamma'\mathbf{\Sigma}_{12}'\boldsymbol{\alpha} - \mu\gamma'\mathbf{\Sigma}_{22}\boldsymbol{\gamma} = 0. \tag{10}$$

由于 $\boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha} = 1$, $\boldsymbol{\gamma}'\mathbf{\Sigma}_{22}\boldsymbol{\gamma} = 1$, 这就得到了 $\lambda = \mu = \boldsymbol{\alpha}'\mathbf{\Sigma}_{12}\boldsymbol{\gamma} = \boldsymbol{\gamma}'\mathbf{\Sigma}_{12}'\boldsymbol{\alpha}$. 由于还有 $\mathbf{\Sigma}_{12}' = \Sigma_{12}$, (7) 式和 (8) 式可写成

$$-\lambda\mathbf{\Sigma}_{11}\boldsymbol{\alpha} + \mathbf{\Sigma}_{12}\boldsymbol{\gamma} = \mathbf{0}, \tag{11}$$

$$\mathbf{\Sigma}_{21}\boldsymbol{\alpha} - \lambda\mathbf{\Sigma}_{22}\boldsymbol{\gamma} = \mathbf{0}, \tag{12}$$

矩阵方程形式为

$$\begin{pmatrix} -\lambda\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & -\lambda\mathbf{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \end{pmatrix} = \mathbf{0}. \tag{13}$$

要使上式有非平凡的解 [解必须满足 (3) 式和 (4) 式], 左端的矩阵必须是奇异的, 即

$$\begin{vmatrix} -\lambda\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & -\lambda\mathbf{\Sigma}_{22} \end{vmatrix} = 0. \tag{14}$$

左端的行列式是一个 p 次多项式. 为了证明这点, 考虑前 p_1 列子式的拉普拉斯展开. 其中一项是 $|-\lambda\mathbf{\Sigma}_{11}| \cdot |-\lambda\mathbf{\Sigma}_{22}| = (-\lambda)^{p_1+p_2}|\mathbf{\Sigma}_{11}| \cdot |\mathbf{\Sigma}_{22}|$. 展开式的其他项是 λ 的低次项, 因为前 p_1 列的子式没有更多的行包含 λ. 由于 $\mathbf{\Sigma}$ 是正定的, $|\mathbf{\Sigma}_{11}| \cdot |\mathbf{\Sigma}_{22}| \neq 0$(见附录的推论 A. 1. 3). 这证明了 (14) 式是一个 p 次多项式方程, 它有 p 个根, 设为 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$. [(9) 式和 (10) 式中的 $\boldsymbol{\alpha}'$ 和 γ' 复共轭证明了 λ 是实数.]

从 (9) 式我们可以得到 $\lambda = \boldsymbol{\alpha}'\mathbf{\Sigma}_{22}\boldsymbol{\gamma}$ 是 $U = \boldsymbol{\alpha}'X^{(1)}$ 和 $V = \gamma'X^{(2)}$ 在 α 和 γ 对某些 λ 值满足 (13) 式的条件下的相关. 我们想找极大相关, 令 $\lambda = \lambda_1$. 对 $\lambda = \lambda_1$, (13) 式的一组解为 $\boldsymbol{\alpha}^{(1)}$, $\boldsymbol{\gamma}^{(1)}$, 令 $U_1 = \boldsymbol{\alpha}^{(1)'}\boldsymbol{X}^{(1)}$, $V_1 = \boldsymbol{\gamma}^{(1)'}\boldsymbol{X}^{(2)}$. U_1 和 V_1 分别是 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的正规化线性组合, 并且它们有极大相关.

我们现在考虑寻找 $\boldsymbol{X}^{(1)}$ 的第二个线性组合, 令 $U = \boldsymbol{\alpha}'\mathbf{X}^{(1)}$, $\boldsymbol{X}^{(2)}$ 的第二个线性组合, 令 $V = \gamma'\boldsymbol{X}^{(2)}$, 它们在所有和 U_1、V_1 不相关的线性组合中有极大相关. 继续这样的步骤. 第 r 步我们得到了线性组合 $U_1 = \boldsymbol{\alpha}^{(1)'}\boldsymbol{X}^{(1)}$, $V_1 = \boldsymbol{\gamma}^{(1)'}\boldsymbol{X}^{(2)}$, $\cdots$, $U_r = \boldsymbol{\alpha}^{(r)'}\boldsymbol{X}^{(1)}$, $V_r = \boldsymbol{\gamma}^{(r)'}\boldsymbol{X}^{(2)}$, 以及对应的相关 [(14) 式的根]$\lambda^{(1)} = \lambda_1$, $\lambda^{(2)}$, $\cdots$, $\lambda^{(r)}$. 我们寻找 $\boldsymbol{X}^{(1)}$ 的一个线性组合 $U = \boldsymbol{\alpha}'\boldsymbol{X}^{(1)}$, $\boldsymbol{X}^{(2)}$ 的线性组合 $V = \gamma'\boldsymbol{X}^{(2)}$, 在与 $U_1, V_1, \cdots, U_r, V_r$ 不相关的所有线性组合中, 有极大相关. 条件 U 和 U_i 不相关, 也即

$$0 = \mathrm{E}(UU_i) = \mathrm{E}(\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}\boldsymbol{X}^{(1)'}\boldsymbol{\alpha}^{(i)}) = \boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha}^{(i)}. \tag{15}$$

则有

$$\mathrm{E}(UV_i) = \boldsymbol{\alpha}'\mathbf{\Sigma}_{12}\boldsymbol{\gamma}^{(i)} = \lambda^{(i)}\boldsymbol{\alpha}'\mathbf{\Sigma}_{11}\boldsymbol{\alpha}^{(i)} = 0. \tag{16}$$

条件 V 和 V_i 不相关, 也即

$$0 = \mathrm{E}(VV_i) = \boldsymbol{\gamma}'\mathbf{\Sigma}_{22}\boldsymbol{\gamma}^{(i)}. \tag{17}$$

由同样的论证我们有

$$\mathrm{E}(VU_i) = \boldsymbol{\gamma}'\mathbf{\Sigma}_{21}\boldsymbol{\alpha}^{(i)} = 0. \tag{18}$$

我们现在要极大化 $\mathrm{E}(U_{r+1}V_{r+1})$, 选择 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 满足 (3), (4), (15) 和 (17), $i=1,2,\cdots,r$. 考虑

$$\psi_{r+1}=\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}-\frac{1}{2}\lambda(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}-1)-\frac{1}{2}\mu(\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}-1)+\sum_{i=1}^{r}\nu_i\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}^{(i)}+\sum_{i=1}^{r}\theta_i\boldsymbol{\gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}^{(i)}, \tag{19}$$

其中 $\lambda,\mu,\nu_1,\cdots,\nu_r,\theta_1,\cdots,\theta_r$ 是拉格朗日乘数. 对 ψ_{r+1} 求关于 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 的偏导数向量, 并令其等于零. 得到

$$\frac{\partial\psi_{r+1}}{\partial\boldsymbol{\alpha}}=\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}-\lambda\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}+\sum_{i=1}^{r}\nu_i\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}^{(i)}=\mathbf{0}, \tag{20}$$

$$\frac{\partial\psi_{r+1}}{\partial\boldsymbol{\gamma}}=\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}-\mu\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}+\sum_{i=1}^{r}\theta_i\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}^{(i)}=\mathbf{0}. \tag{21}$$

(20) 式左端乘 $\boldsymbol{\alpha}^{(j)\prime}$, (21) 式左端乘 $\boldsymbol{\gamma}^{(j)\prime}$, 得到

$$0=\nu_j\boldsymbol{\alpha}^{(j)\prime}\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}^{(j)}=\nu_j, \tag{22}$$

$$0=\theta_j\boldsymbol{\gamma}^{(j)\prime}\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}^{(j)}=\theta_j. \tag{23}$$

因此 (20) 和 (21) 就简化为 (11) 和 (12) 或者替代的 (13). 我们取最大的 λ_i, 并设为 $\lambda^{(r+1)}$, 因此 (13) 式有一组解, 满足 (3), (4), (15) 和 (17), $i=1,\cdots,r$. 令这组解为 $\boldsymbol{\alpha}^{(r+1)}$, $\boldsymbol{\gamma}^{(r+1)}$, 令 $U_{r+1}=\boldsymbol{\alpha}^{(r+1)\prime}\boldsymbol{X}^{(1)}$, $V_{r+1}=\boldsymbol{\gamma}^{(r+1)\prime}\boldsymbol{X}^{(2)}$.

这样的步骤一步一步继续直到找到满足条件的相继解, 也就是 (13) 对某些 λ_i, 满足 (3) , (4), (15) 和 (17). 设这样的步骤进行了 m 步. 我们现在证明 $m=p_1(\leqslant p_2)$. 令 $\boldsymbol{A}=(\boldsymbol{\alpha}^{(1)},\cdots,\boldsymbol{\alpha}^{(m)}),\boldsymbol{\Gamma}_1=(\boldsymbol{\gamma}^{(1)},\cdots,\boldsymbol{\gamma}^{(m)})$, 还有

$$\boldsymbol{\Lambda}=\begin{pmatrix}\lambda^{(1)} & 0 & \cdots & 0\\ 0 & \lambda^{(2)} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda^{(m)}\end{pmatrix}. \tag{24}$$

条件 (3) 和 (15) 可以概括成

$$\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{A}=\boldsymbol{I}. \tag{25}$$

由于 $\boldsymbol{\Sigma}_{11}$ 的秩是 p_1, $\boldsymbol{I}$ 的秩是 m, 我们有 $m\leqslant p_1$. 现在我们证明若 $m<p_1$, 则存在另一组满足条件的向量, 导致了矛盾. 由于 $\boldsymbol{A}'\boldsymbol{\Sigma}_{11}$ 是 $m\times p_1$ 维的, 存在一个 $p_1\times(p_1-m)$ 维矩阵 $\boldsymbol{E}$(秩为 p_1-m), 使得 $\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}=\mathbf{0}$. 类似地, 存在一个 $p_2\times(p_2-m)$ 维矩阵 $\boldsymbol{F}$(秩为 p_2-m), 使得 $\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{22}\boldsymbol{F}=\mathbf{0}$. 同样还有 $\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{21}\boldsymbol{E}=\boldsymbol{\Lambda}\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}=\mathbf{0}$, $\boldsymbol{A}'\boldsymbol{\Sigma}_{12}\boldsymbol{F}=\boldsymbol{\Lambda}\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{22}\boldsymbol{F}=\mathbf{0}$. 由于 $\boldsymbol{E}$ 的秩是 p_1-m, 故 $\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}$ 是非奇异的 (若 $m<p_1$), 类似地, $\boldsymbol{F}'\boldsymbol{\Sigma}_{22}\boldsymbol{F}$ 也是非奇异的. 因此

$$\begin{vmatrix}-\nu\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E} & \boldsymbol{E}'\boldsymbol{\Sigma}_{12}\boldsymbol{F}\\ \boldsymbol{F}'\boldsymbol{\Sigma}_{21}\boldsymbol{E} & -\nu\boldsymbol{F}'\boldsymbol{\Sigma}_{22}\boldsymbol{F}\end{vmatrix}=0 \tag{26}$$

至少有一个根, 这是因为 $|\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}|\cdot|\boldsymbol{F}'\boldsymbol{\Sigma}_{22}\boldsymbol{F}|\neq 0$. 从上式我们可以看出, 存在向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 满足

$$\boldsymbol{E}'\boldsymbol{\Sigma}_{12}\boldsymbol{F}\boldsymbol{b}=\nu\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}\boldsymbol{a}, \tag{27}$$

$$\boldsymbol{F}'\boldsymbol{\Sigma}_{21}\boldsymbol{E}\boldsymbol{a}=\nu\boldsymbol{F}'\boldsymbol{\Sigma}_{22}\boldsymbol{F}\boldsymbol{b}. \tag{28}$$

令 $\boldsymbol{E}\boldsymbol{a}=\boldsymbol{g}$, $\boldsymbol{F}\boldsymbol{b}=\boldsymbol{h}$. 现在我们想要证明 $\boldsymbol{\nu},\boldsymbol{g},\boldsymbol{h}$ 是一组新的解 $\lambda^{(m+1)},\boldsymbol{\alpha}^{(m+1)}$, $\gamma^{(m+1)}$. 令 $\boldsymbol{\Sigma}_{11}{}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{h}=\boldsymbol{k}$. 因为 $\mathbf{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{k}=\mathbf{A}'\boldsymbol{\Sigma}_{12}\boldsymbol{F}\boldsymbol{b}=\mathbf{0}$, $\boldsymbol{k}$ 和 $\boldsymbol{A}'\boldsymbol{\Sigma}_{11}$ 的行向量正交, 故存在 $\boldsymbol{E}$ 的列向量的线性组合, 设为 $\boldsymbol{E}\boldsymbol{c}$. 因此等式 $\boldsymbol{\Sigma}_{12}\boldsymbol{h}=\boldsymbol{\Sigma}_{11}\boldsymbol{k}$ 可以写成

$$\boldsymbol{\Sigma}_{12}\boldsymbol{F}\boldsymbol{b}=\boldsymbol{\Sigma}_{11}\boldsymbol{E}\boldsymbol{c}. \tag{29}$$

左端乘 E' 得到

$$\boldsymbol{E}'\boldsymbol{\Sigma}_{12}\boldsymbol{F}\boldsymbol{b}=\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}\boldsymbol{c}. \tag{30}$$

由于 $\boldsymbol{E}'\boldsymbol{\Sigma}_{11}\boldsymbol{E}$ 是非奇异的. 比较 (27) 和 (30) 可得 $\boldsymbol{c}=\boldsymbol{\nu}\boldsymbol{a}$, 因此 $\boldsymbol{k}=\boldsymbol{\nu}\boldsymbol{g}$. 于是

$$\boldsymbol{\Sigma}_{12}\boldsymbol{h}=\nu\boldsymbol{\Sigma}_{11}g. \tag{31}$$

由相同的方式我们得到

$$\boldsymbol{\Sigma}_{21}g=\nu\boldsymbol{\Sigma}_{22}h. \tag{32}$$

因此 $\nu=\lambda^{(m+1)},g=\boldsymbol{\alpha}^{(m+1)},h=\gamma^{(m+1)}$ 是另一组解. 但是这和我们的假设 $\lambda^{(m)},\boldsymbol{\alpha}^{(m)},\gamma^{(m)}$ 是最后一组可能的解矛盾. 因此有 $m=p_1$.

关于 $\lambda,\boldsymbol{\alpha},\gamma$ 的这些条件可以概括成

$$\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{A}=\boldsymbol{I}, \tag{33}$$

$$\boldsymbol{A}'\boldsymbol{\Sigma}_{12}\boldsymbol{\Gamma}_1=\boldsymbol{\Lambda}, \tag{34}$$

$$\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}_{22}\boldsymbol{\Gamma}_1=\boldsymbol{I}. \tag{35}$$

令 $\boldsymbol{\Gamma}_2=(\gamma^{(p_1+1)}\cdots\gamma^{(p_2)})$ 是 $p_2\times(p_2-p_1)$ 矩阵, 满足

$$\boldsymbol{\Gamma}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\Gamma}_1=\mathbf{0}, \tag{36}$$

$$\boldsymbol{\Gamma}_2'\boldsymbol{\Sigma}_{22}\boldsymbol{\Gamma}_2=\boldsymbol{I}. \tag{37}$$

任意 $\boldsymbol{\Gamma}_2$ 都可以右乘任意一个 $(p_2-p_1)\times(p_2-p_1)$ 正交矩阵. 这个矩阵一次可以形成一列; $\gamma^{(p_1+1)}$ 是与 $\boldsymbol{\Sigma}_{22}\boldsymbol{\Gamma}_1$ 正交的向量, 正规化有 $\gamma^{(p_1+1)'}\boldsymbol{\Sigma}_{22}\gamma^{(p_1+1)}=1$; $\gamma^{(p_1+2)}$ 是与 $\boldsymbol{\Sigma}_{22}(\boldsymbol{\Gamma}_1\quad\gamma^{(p_1+1)})$ 正交的向量, 将其正规化有 $\gamma^{(p_1+2)'}\boldsymbol{\Sigma}_{22}\gamma^{(p_1+2)}=1$; 依此类推. 令 $\boldsymbol{\Gamma}=(\boldsymbol{\Gamma}_1\ \boldsymbol{\Gamma}_2)$, 因为 $\boldsymbol{\Gamma}'\boldsymbol{\Sigma}_{22}\boldsymbol{\Gamma}=\boldsymbol{I}$, 所以矩阵 $\boldsymbol{\Gamma}$ 是非奇异的. 考虑行列式

$$\begin{vmatrix}\boldsymbol{A}' & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Gamma}_1'\\ \mathbf{0} & \boldsymbol{\Gamma}_2'\end{vmatrix}\cdot\begin{vmatrix}-\lambda\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & -\lambda\boldsymbol{\Sigma}_{22}\end{vmatrix}\cdot\begin{vmatrix}\boldsymbol{A} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2\end{vmatrix} \tag{38}$$

$$=\begin{vmatrix}-\lambda\boldsymbol{I} & \boldsymbol{\Lambda} & \mathbf{0}\\ \boldsymbol{\Lambda} & -\lambda\boldsymbol{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & -\lambda\boldsymbol{I}\end{vmatrix}$$

$$
\begin{aligned}
&=(-\lambda)^{(p_2-p_1)}\begin{vmatrix} -\lambda \boldsymbol{I} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda} & -\lambda \boldsymbol{I} \end{vmatrix} \\
&=(-\lambda)^{(p_2-p_1)}|-\lambda \boldsymbol{I}|\cdot|-\lambda \boldsymbol{I}-\boldsymbol{\Lambda}(-\lambda \boldsymbol{I})^{-1}\boldsymbol{\Lambda}| \\
&=(-\lambda)^{(p_2-p_1)}|\lambda^2 \boldsymbol{I}-\boldsymbol{\Lambda}^2| \\
&=(-\lambda)^{(p_2-p_1)}\prod(\lambda^2-\lambda^{(i)2}).
\end{aligned}
$$

除了常数因子, 上面的多项式是

$$
\begin{vmatrix} -\lambda \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda \boldsymbol{\Sigma}_{22} \end{vmatrix}. \tag{39}
$$

因此 (14) 式的根正是令 (38) 式等于零的根, 即 $\lambda = \pm\lambda^{(i)}$, $i = 1,\cdots,p_1$, 以及 $\lambda = 0$(有重数 $p_2 - p_1$). 因此 $(\lambda_1,\cdots,\lambda_p) = (\lambda_1,\cdots,\lambda_{p_1},0,\cdots,0,-\lambda_{p_1},\cdots,-\lambda_1)$. 集合 $\{\lambda^{(i)2}\}$, $i = 1,\cdots,p_1$, 也就是集合 $\{\lambda_i{}^2\}$, $i = 1,\cdots,p_1$. 为了证明集合 $\{\lambda^{(i)}\}$, $i = 1,\cdots,p_1$, 就是集合 $\{\lambda_i\}$, $i = 1,\cdots,p_1$, 我们只需要证明 $\lambda^{(i)}$ 是非负的 (因此是 λ_i 中的某一个, $i = 1,\cdots,p_1$). 我们观察到

$$
\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}^{(r)} = -\lambda^{(r)}\boldsymbol{\Sigma}_{11}(-\boldsymbol{\alpha}^{(r)}), \tag{40}
$$

$$
\boldsymbol{\Sigma}_{21}(-\boldsymbol{\alpha}^{(r)}) = -\lambda^{(r)}\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}^{(r)}; \tag{41}
$$

因此, 如果 $\lambda^{(r)},\boldsymbol{\alpha}^{(r)},\boldsymbol{\gamma}^{(r)}$ 是一组解, 那么 $-\lambda^{(r)},-\boldsymbol{\alpha}^{(r)},-\boldsymbol{\gamma}^{(r)}$ 也是一组解. 如果 $\lambda^{(r)}$ 是负的, $-\lambda^{(r)}$ 则是正的, 有 $-\lambda^{(r)} \geqslant \lambda^{(r)}$. 但是由于已知 $\lambda^{(r)}$ 是最大的, 所以必须有 $\lambda^{(r)} \geqslant -\lambda^{(r)}$. 因此 $\lambda^{(r)} \geqslant 0$. 由于集合 $\{\lambda^{(i)}\}$, $i = 1,\cdots,p_1$, 就是集合 $\{\lambda_i\}$, $i = 1,\cdots,p_1$, 所以必须有 $\lambda^{(i)} = \lambda_i$.

令

$$
\boldsymbol{U} = \begin{pmatrix} U_1 \\ \vdots \\ U_{p_1} \end{pmatrix} = \boldsymbol{A}'\boldsymbol{X}^{(1)}, \tag{42}
$$

$$
\boldsymbol{V}^{(1)} = \begin{pmatrix} V_1 \\ \vdots \\ V_{p_1} \end{pmatrix} = \boldsymbol{\Gamma}_1'\boldsymbol{X}^{(2)}, \tag{43}
$$

$$
\boldsymbol{V}^{(2)} = \begin{pmatrix} V_{p_1+1} \\ \vdots \\ V_{p_2} \end{pmatrix} = \boldsymbol{\Gamma}_2'\boldsymbol{X}^{(2)}. \tag{44}
$$

$\boldsymbol{U}$ 的分量是一组典型变量, $\boldsymbol{V} = (\boldsymbol{V}^{(1)\prime}\ \boldsymbol{V}^{(2)\prime})'$ 的分量是另一组典型变量. 我们有

$$\mathrm{E}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V}^{(1)} \\ \boldsymbol{V}^{(2)} \end{pmatrix}\begin{pmatrix} \boldsymbol{U}' & \boldsymbol{V}^{(1)'} & \boldsymbol{V}^{(2)'} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma}_1' \\ \boldsymbol{0} & \boldsymbol{\Gamma}_2' \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{I}_{p_1} & \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{\Lambda} & \boldsymbol{I}_{p_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{p_2-p_1} \end{pmatrix}, \tag{45}$$

其中

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_{p_1} \end{pmatrix}. \tag{46}$$

定义 12.2.1 令 $\boldsymbol{X} = (\boldsymbol{X}^{(1)'} \quad \boldsymbol{X}^{(2)'})'$, 其中 $\boldsymbol{X}^{(1)}$ 是 p_1 元向量, $\boldsymbol{X}^{(2)}$ 是 $p_2(=p-p_1 \geqslant p_1)$ 元向量. 第 r 组典型变量是第 r 组线性组合 $U_r = \boldsymbol{\alpha}^{(r)'}\boldsymbol{X}^{(1)}, \boldsymbol{V}_r = \boldsymbol{\gamma}^{(r)'}\boldsymbol{X}^{(2)}$, 每组典型变量都有单位方差, 它们和前 $r-1$ 组典型变量是不相关的, 并且它们之间有极大相关. 它们的相关是第 r 组典型相关.

定理 12.2.1 令 $\boldsymbol{X} = (\boldsymbol{X}^{(1)'} \quad \boldsymbol{X}^{(2)'})$ 是随机向量, 协方差阵是 $\boldsymbol{\Sigma}$. $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 的第 r 组典型相关是 (14) 式的第 r 个最大根. $\boldsymbol{\alpha}^{(r)'}\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{\gamma}^{(r)'}\boldsymbol{X}^{(2)}$ 的系数定义了第 r 组典型变量, 对 $\lambda = \lambda_r$ 满足 (13) 式、(3) 式和 (4) 式.

我们现在可以证明 (不用微分法)U_1 和 V_1 有极大相关. 线性组合 $\boldsymbol{a}'\boldsymbol{U} = (\boldsymbol{a}'\boldsymbol{A}')\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{b}'\boldsymbol{V} = (\boldsymbol{b}'\boldsymbol{\Gamma}')\boldsymbol{X}^{(2)}$ 由 $\boldsymbol{a}'\boldsymbol{a} = 1$ 和 $\boldsymbol{b}'\boldsymbol{b} = 1$ 正规化. 由于 $\boldsymbol{A}$ 和 $\boldsymbol{\Gamma}$ 是非奇异的, 任意向量 $\boldsymbol{\alpha}$ 可写成 $\boldsymbol{A}\boldsymbol{a}$, 任意向量 $\boldsymbol{\gamma}$ 可写成 $\boldsymbol{\Gamma}\boldsymbol{b}$, 因此任意线性组合 $\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{\gamma}'\boldsymbol{X}^{(2)}$ 可写成 $\boldsymbol{a}'\boldsymbol{U}$ 和 $\boldsymbol{b}'\boldsymbol{V}$. 它们之间的相关是

$$\boldsymbol{a}'\begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \end{pmatrix}\boldsymbol{b} = \sum_{i=1}^{p_1} \lambda_i a_i b_i. \tag{47}$$

令 $\lambda_i a_i / \sqrt{\sum(\lambda_i a_i)^2} = c_i$. 则

$$\boldsymbol{a}'\begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \end{pmatrix}\boldsymbol{b} = \sqrt{\sum(\lambda_i a_i)^2}\sum c_i b_i$$

关于 b 的极大值是取 $b_i = c_i$, 因为 $\sum c_i b_i$ 是向量 b 和 $(c_1, \cdots, c_{p_1}, 0, \cdots, 0)$ 夹角的余弦. 则 (47) 式即

$$\sqrt{\sum \lambda_i^2 a_i^2} = \sqrt{\sum\nolimits_2^{p_1} (\lambda_i^2 - \lambda_1^2)a_i^2 + \lambda_1^2},$$

当 $a_i = 0(i = 2, \cdots, p_1)$ 时上式取极大值. 因此极大化线性组合是 U_1 和 V_1. 为了证明 U_2 和 V_2 是第二组典型变量, 我们注意到 U_1 和线性组合 $\boldsymbol{a}'\boldsymbol{U}$ 缺少相关意味着 $0 = \mathrm{E}(\boldsymbol{U}_1\boldsymbol{a}'\boldsymbol{U}) = \mathrm{E}(U_1\Sigma a_i U_i) = a_1$, V_1 和 $\boldsymbol{b}'\boldsymbol{V}$ 缺少相关意味着 $0 = b_1$. 上面的代数式给出了从 $i = 2$ 开始求和的相应的结果.

对 α 或 γ, 我们可以得到单个矩阵方程. 如果把 (11) 式乘 λ , (12) 式乘 $\boldsymbol{\Sigma}_{22}{}^{-1}$, 我们得到

$$\lambda\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}=\lambda^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}, \tag{48}$$

$$\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}=\lambda\boldsymbol{\gamma}. \tag{49}$$

将 (49) 式代入 (48) 式得到

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}=\lambda^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} \tag{50}$$

或者

$$(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\lambda^2\boldsymbol{\Sigma}_{11})\boldsymbol{\alpha}=\mathbf{0}. \tag{51}$$

$\lambda_1{}^2,\cdots,\lambda_{p_1}{}^2$ 满足

$$|\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\nu\boldsymbol{\Sigma}_{11}|=0, \tag{52}$$

$\boldsymbol{\alpha}^{(1)},\cdots,\boldsymbol{\alpha}^{(p_1)}$ 分别对 $\lambda^2=\lambda_1{}^2,\cdots,\lambda_{p_1}{}^2$ 满足 (51). 类似地, 当 $\lambda^2=\lambda_1{}^2,\cdots,\lambda_{p_2}{}^2$ 时关于 $\gamma^{(1)},\cdots,\gamma^{(p_2)}$ 的方程被

$$(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}-\lambda^2\boldsymbol{\Sigma}_{22})\boldsymbol{\gamma}=\mathbf{0} \tag{53}$$

替代.

定理 12.2.2　*典型相关对于变换 $\mathbf{X}^{(i)*}=\boldsymbol{C}_i\mathbf{X}^i$ 是不变量, 其中 $\boldsymbol{C}_i$ 是非奇异的, $i=1,2$, 如果 $\boldsymbol{\Sigma}$ 的任意函数是不变量, 那么它是典型相关的函数.*

证明　(14) 式变换为

$$\begin{aligned}0&=\begin{vmatrix}-\lambda\boldsymbol{C}_1\boldsymbol{\Sigma}_{11}\boldsymbol{C}_1{}' & \boldsymbol{C}_1\boldsymbol{\Sigma}_{12}\boldsymbol{C}_2{}'\\ \boldsymbol{C}_2\boldsymbol{\Sigma}_{21}\boldsymbol{C}_1{}' & -\lambda\boldsymbol{C}_2\boldsymbol{\Sigma}_{22}\boldsymbol{C}_2{}'\end{vmatrix}\\ &=\begin{vmatrix}\boldsymbol{C}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}_2\end{vmatrix}\cdot\begin{vmatrix}-\lambda\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & -\lambda\boldsymbol{\Sigma}_{22}\end{vmatrix}\cdot\begin{vmatrix}\boldsymbol{C}_1{}' & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}_2{}'\end{vmatrix},\end{aligned} \tag{54}$$

因此根没有变. 相反地, 令 $f(\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{12},\boldsymbol{\Sigma}_{22})$ 是 $\boldsymbol{\Sigma}$ 的向量值函数, 使得对所有非奇异的 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$, $f(\boldsymbol{C}_1\boldsymbol{\Sigma}_{11}\boldsymbol{C}_1{}',\boldsymbol{C}_1\boldsymbol{\Sigma}_{12}\boldsymbol{C}_2{}',\boldsymbol{C}_2\boldsymbol{\Sigma}_{22}\boldsymbol{C}_2{}')=f(\boldsymbol{\Sigma}_{11},\boldsymbol{\Sigma}_{12},\boldsymbol{\Sigma}_{22})$. 如果 $\boldsymbol{C}_1{}'=\boldsymbol{A}$, $\boldsymbol{C}_2=\boldsymbol{\Gamma}'$, 那么 (54) 式也即 (38) 式, 它只取决于典型相关. 故 $f=f(\boldsymbol{I},(\boldsymbol{\Lambda},\mathbf{0}),\boldsymbol{I})$.■

我们对这些展开式用预报的观点做另外一种解释. 考虑两个随机变量 U 和 V, 它们均值为 0, 方差分别是 $\sigma_u{}^2$ 和 $\sigma_v{}^2$, 它们之间的相关是 ρ. 考虑用 V 的倍数近似 U, 设为 bV; 则近似值的均方误差是

$$\begin{aligned}\mathrm{E}(U-bV)^2&=\sigma_u{}^2-2b\sigma_u\sigma_v\rho+b^2\sigma_v{}^2\\ &=\sigma_u{}^2(1-\rho^2)+(b\sigma_v-\rho\sigma_u)^2.\end{aligned} \tag{55}$$

当 $b=\sigma_u\rho/\sigma_v$ 时上式达到极小. 我们可将 bV 视为由 V 对 U 的线性预报, 则 $\sigma_u{}^2(1-\rho^2)$ 是预报的均方误差. 预报的均方误差和 U 的方差的比是 $\sigma_u{}^2(1-\rho^2)/\sigma_u{}^2=1-\rho^2$, 余集是 V 对 U 的相对效应的度量或是 V 对 U 的预报的有效性的度量. 因此 ρ^2 或者 $|\rho|$ 越大, V 对 U 的预报越有效.

现在我们考虑随机向量 $\boldsymbol{X}$ 按照 (1) 式的划分, 考虑用一个线性组合 $V = \boldsymbol{\gamma}'\boldsymbol{X}^{(2)}$ 来预报一个线性组合 $U = \boldsymbol{\alpha}'\boldsymbol{X}^{(1)}$. 如果 U 和 V 的相关达到极大, 则 V 对 U 的预报达到最优. 因此我们可以说 $\boldsymbol{\alpha}^{(1)\prime}\boldsymbol{X}^{(1)}$ 是 $\boldsymbol{X}^{(1)}$ 的能达到最优预报的线性组合, $\boldsymbol{\gamma}^{(1)\prime}\boldsymbol{X}^{(2)}$ 就是最优预报值 [Hotelling(1935)].

V 对 U 的均方效应可以用

$$\mathrm{E}(bV)^2 = \rho^2 \frac{{\sigma_u}^2}{{\sigma_v}^2}\mathrm{E}(V^2) = \rho^2{\sigma_u}^2 \tag{56}$$

来度量, 相应的均方误差效应可以用 $\mathrm{E}(bV)^2/\mathrm{E}(U^2) = \rho^2$ 来度量. 因此 $\boldsymbol{X}^{(2)}$ 的线性组合对 $\boldsymbol{X}^{(1)}$ 的线性组合的极大效应是 $\boldsymbol{\gamma}^{(1)\prime}\boldsymbol{X}^{(2)}$ 对 $\boldsymbol{\alpha}^{(1)\prime}\boldsymbol{X}^{(1)}$.

对 $p_1 = 1$ 的特殊情况, 一个典型相关是 $\boldsymbol{X}^{(1)} = X_1$ 和 $\boldsymbol{X}^{(2)}$ 之间的多重相关系数.

典型变量和典型相关的定义是用协方差阵 $\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}))(\boldsymbol{X} - \mathrm{E}(\boldsymbol{X}))'$ 的方法. 我们可以从一个 $p + p_3$ 元的正态分布的向量 $\boldsymbol{Y}$ 出发来拓展这个定义, 定义 $\boldsymbol{X}$ 为一个向量, 它的分布是, 当给定 $\boldsymbol{Y}$ 的后 p_3 个分量的值时前 p 个分量的条件分布. 这意味着对 $\boldsymbol{X}_\phi$ 有均值 $\mathrm{E}(\boldsymbol{X}_\phi) = \boldsymbol{\Theta}\boldsymbol{y}_\phi^{(3)}$, 协方差阵的元素是 $\boldsymbol{Y}$ 的前 P 个分量的偏协方差.

典型向量的解释在考虑典型变量和初始向量的分量之间的相关下得到简化 [例如, Darlington, Weinberg, and Wahlberg(1973)]. 第 j 个典型变量 U_j 和 X_i 的协方差是

$$\mathrm{E}(U_jX_i) = \mathrm{E}\sum_{k=1}^{p_1}{\alpha_k}^{(j)}X_kX_i = \sum_{k=1}^{p_1}{\alpha_k}^{(j)}\sigma_{ki}. \tag{57}$$

由于 U_j 的方差是 1, U_j 和 X_i 的相关是

$$\mathrm{Corr}(U_j, X_i) = \frac{\sum_{k=1}^{p_1}{\alpha_k}^{(j)}\sigma_{ki}}{\sqrt{\sigma_{ii}}}. \tag{58}$$

这种度量的一个优点是它不依赖于 X_j 的测量单位. 尽管如此, 但它不是 X_i 对 U_j(即 ${\alpha_i}^{(j)}$) 的权的标量倍数.

一个特殊的情况是 $\boldsymbol{\Sigma}_{11} = \boldsymbol{I}, \boldsymbol{\Sigma}_{22} = \boldsymbol{I}$. 则

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}, \quad \boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}, \quad \mathbf{A}'\boldsymbol{\Sigma}_{12}\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \end{pmatrix}. \tag{59}$$

从这些我们得到

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{A}\begin{pmatrix} \boldsymbol{\Lambda} & \mathbf{0} \end{pmatrix}\boldsymbol{\Gamma}', \tag{60}$$

其中 $\boldsymbol{A}$ 和 $\boldsymbol{\Gamma}$ 是正交的, $\boldsymbol{\Lambda}$ 是对角矩阵. 这个关系由 $\boldsymbol{\Sigma}_{12}$ 的奇异值分解得到. $\boldsymbol{\Lambda}$ 的元是 $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{12}'$ 的特征根的平方根, $\boldsymbol{A}$ 的列是其特征向量. $\boldsymbol{\Lambda}$ 的对角元素是 $\boldsymbol{\Sigma}_{12}'\boldsymbol{\Sigma}_{12}$ 的特征根(可能非零) 的平方根, $\boldsymbol{\Gamma}$ 的列是其特征向量.

12.3　典型相关和典型变量的估计

12.3.1　估计

令 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 N 个观测. 令 $\boldsymbol{x}_\alpha$ 分为 p_1 元和 p_2 元两部分,

$$\boldsymbol{x}_\alpha = \begin{pmatrix} \boldsymbol{x}_\alpha^{(1)} \\ \boldsymbol{x}_\alpha^{(2)} \end{pmatrix}, \quad \alpha = 1, \cdots, N. \tag{1}$$

$\boldsymbol{\Sigma}$ 的极大似然估计量是 [按 12.2 节中 (2) 式划分]

$$\begin{aligned}\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} \end{pmatrix} &= \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' \\ &= \frac{1}{N}\begin{pmatrix} \sum(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})' & \sum(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})' \\ \sum(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})' & \sum(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})(\boldsymbol{x}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})' \end{pmatrix}.\end{aligned} \tag{2}$$

典型相关 $\boldsymbol{\Lambda}$ 和由 $\mathbf{A}$ 和 $\boldsymbol{\Gamma}$ 定义的典型变量的极大似然估计量, 包括把前面的代数知识应用到 $\hat{\boldsymbol{\Sigma}}$. 如果假定典型相关是不同的, $\mathbf{A}$ 的每列非零元素均为正的, 则矩阵 $\boldsymbol{\Lambda}, \mathbf{A}, \boldsymbol{\Gamma}_1$ 是唯一定义的. $\boldsymbol{\Gamma}_2$ 的不确定性允许在右端乘一个 $(p_2 - p_1) \times (p_2 - p_1)$ 正交矩阵. 这种不确定性可通过增加不同的限制消除, 比如, 由下面 $p_2 - p_1$ 行构成的子矩阵是上三角的或下三角的, 且对角元为正. 应用推论 3.2.1 得到 $\lambda_1, \cdots, \lambda_p$ 的极大似然估计量是

$$\begin{vmatrix} -l\hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & -l\hat{\boldsymbol{\Sigma}}_{22} \end{vmatrix} = 0 \tag{3}$$

的根. $\hat{\mathbf{A}}$ 和 $\hat{\boldsymbol{\Gamma}}_1$ 的第 j 列满足

$$\begin{pmatrix} -l_j\hat{\boldsymbol{\Sigma}}_{11} & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & -l_j\hat{\boldsymbol{\Sigma}}_{22} \end{pmatrix}\begin{pmatrix} \hat{\boldsymbol{\alpha}}^{(j)} \\ \hat{\boldsymbol{\gamma}}^{(j)} \end{pmatrix} = \mathbf{0}, \tag{4}$$

$$\hat{\boldsymbol{\alpha}}^{(j)\prime}\hat{\boldsymbol{\Sigma}}_{11}\hat{\boldsymbol{\alpha}}^{(j)} = 1, \qquad \hat{\boldsymbol{\gamma}}^{(j)\prime}\hat{\boldsymbol{\Sigma}}_{22}\hat{\boldsymbol{\gamma}}^{(j)} = 1. \tag{5}$$

$\hat{\boldsymbol{\Gamma}}_2$ 满足

$$\hat{\boldsymbol{\Gamma}}_2'\hat{\boldsymbol{\Sigma}}_{22}\hat{\boldsymbol{\Gamma}}_1 = \mathbf{0}, \tag{6}$$

$$\hat{\boldsymbol{\Gamma}}_2'\hat{\boldsymbol{\Sigma}}_{22}\hat{\boldsymbol{\Gamma}}_2 = \boldsymbol{I}. \tag{7}$$

当对 $\boldsymbol{\Gamma}_2$ 加上另外的限制时, 则 $\hat{\mathbf{A}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Lambda}}$ 就唯一定义了.

定理 12.3.1　*令 $x_1, \cdots, x_N$ 是来自 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 N 个观测. 令 $\boldsymbol{\Sigma}$ 如 12.2 节中 (2) 式的将行和列划分为 p_1 和 $p_2(p_1 \leqslant p_2)$ 部分, 令 x_α 有类似 (1) 式的划分. 典型相关的极大似然估计量是 (3) 的根, 其中 $\hat{\boldsymbol{\Sigma}}_{ij}$ 是由 (2) 定义的. 第 j 组典型分量的*

系数的极大似然估计量满足 (4) 式和 (5) 式, $j=1,\cdots,p_1$; 其余的分量满足 (6) 式和 (7) 式.

总体的典型相关和典型变量是用极大化两组变量的线性组合之间的相关的方法来得到的. 整个讨论可以用于样本. 因而, $\hat{\boldsymbol{\alpha}}^{(1)\prime}\boldsymbol{x}_\alpha^{(1)}$ 和 $\hat{\boldsymbol{\gamma}}^{(1)\prime}\boldsymbol{x}_\alpha^{(2)}$ 在 $\boldsymbol{x}_\alpha^{(1)}$ 和 $\boldsymbol{x}_\alpha^{(2)}$ 的任意线性组合中有极大样本相关, 其相关是 l_1. 类似地, $\hat{\boldsymbol{\alpha}}^{(2)\prime}\boldsymbol{x}_\alpha^{(1)}$ 和 $\hat{\boldsymbol{\gamma}}^{(2)\prime}\boldsymbol{x}_\alpha^{(2)}$ 有第二个极大样本相关, 如此等等.

我们可以用 $\boldsymbol{S}$, 即 $\boldsymbol{\Sigma}$ 的无偏估计量来定义样本的典型变量和典型相关. 则 $\boldsymbol{a}^{(j)}=\sqrt{(N-1)/N}\hat{\boldsymbol{\alpha}}^{(j)}, \boldsymbol{c}^{(j)}=\sqrt{(N-1)/N}\hat{\boldsymbol{\gamma}}^{(j)}$, l_j 满足

$$\boldsymbol{S}_{12}\boldsymbol{c}^{(j)}=l_j\boldsymbol{S}_{11}\boldsymbol{a}^{(j)}, \tag{8}$$

$$\boldsymbol{S}_{21}\boldsymbol{a}^{(j)}=l_j\boldsymbol{S}_{22}\boldsymbol{c}^{(j)}, \tag{9}$$

$$\boldsymbol{a}^{(j)\prime}\boldsymbol{S}_{11}\boldsymbol{a}^{(j)}=1, \qquad \boldsymbol{c}^{(j)\prime}\boldsymbol{S}_{22}\boldsymbol{c}^{(j)}=1. \tag{10}$$

我们将线性组合 $\boldsymbol{a}^{(j)\prime}\boldsymbol{x}_\alpha^{(1)}$ 和 $\boldsymbol{c}^{(j)\prime}\boldsymbol{x}_\alpha^{(2)}$ 称为样本典型变量.

我们可以从样本相关矩阵得到样本典型变量,

$$\boldsymbol{R}=\left(\frac{\hat{\boldsymbol{\sigma}}_{ij}}{\sqrt{\hat{\boldsymbol{\sigma}}_{ii}}\sqrt{\hat{\boldsymbol{\sigma}}_{jj}}}\right)=\left(\frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}\right)=(\boldsymbol{r}_{ij})=\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{pmatrix}. \tag{11}$$

令

$$\boldsymbol{S}_1=\begin{pmatrix}\sqrt{s_{11}} & 0 & \cdots & 0\\ 0 & \sqrt{s_{22}} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sqrt{s_{p_1p_1}}\end{pmatrix}, \tag{12}$$

$$\boldsymbol{S}_2=\begin{pmatrix}\sqrt{s_{p_1+1,p_1+1}} & 0 & \cdots & 0\\ 0 & \sqrt{s_{p_1+2,p_1+2}} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sqrt{s_{pp}}\end{pmatrix}. \tag{13}$$

然后我们根据 (10) 式将 (8) 式写成

$$\boldsymbol{R}_{12}(\boldsymbol{S}_2\boldsymbol{c}^{(j)})=l_j\boldsymbol{R}_{11}(\boldsymbol{S}_1\boldsymbol{a}^{(j)}), \tag{14}$$

$$\boldsymbol{R}_{21}(\boldsymbol{S}_1\boldsymbol{a}^{(j)})=l_j\boldsymbol{R}_{22}(\boldsymbol{S}_2\boldsymbol{c}^{(j)}), \tag{15}$$

$$(\boldsymbol{S}_1\boldsymbol{a}^{(j)})'\boldsymbol{R}_{11}(\boldsymbol{S}_1\boldsymbol{a}^{(j)})=1, \qquad (\boldsymbol{S}_2\boldsymbol{c}^{(j)})'\boldsymbol{R}_{22}(\boldsymbol{S}_2\mathrm{c}^{(j)})=1. \tag{16}$$

我们可以给这些结果一个几何的解释. 矩阵 $(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)$ 的行可以解释为一个 N 维空间中的 p 元向量, $(\boldsymbol{x}_1-\bar{\boldsymbol{x}},\cdots,\boldsymbol{x}_N-\bar{\boldsymbol{x}})$ 的行是投影到和等角线正交的 $(N-1)$ 维子空间上的 p 元向量. 将它们表示为 $\boldsymbol{x}_1^*,\cdots,\boldsymbol{x}_p^*$. 向量 $\boldsymbol{u}^*$ 在 $\boldsymbol{x}_1^*,\cdots,\boldsymbol{x}_{p_1}^*$ 张成的 p_1 维子空间中, 它有分量 $\alpha'(\boldsymbol{x}_1^{(1)}-\bar{\boldsymbol{x}}^{(1)},\cdots,\boldsymbol{x}_N^{(1)}-\bar{\boldsymbol{x}}^{(1)})=\alpha_1\boldsymbol{x}_1^*+\cdots+\alpha_{p_1}\boldsymbol{x}_{p_1}^*$, 向量 $\boldsymbol{v}^*$ 在 $\boldsymbol{x}_{p_1+1}^*,\cdots,\boldsymbol{x}_p^*$ 张成的 p_2 维子空间中, 它有分量 $\gamma'(\boldsymbol{x}_1^{(2)}-\bar{\boldsymbol{x}}^{(2)},\cdots,\boldsymbol{x}_N^{(2)}-$

$\bar{\boldsymbol{x}}^{(2)}) = \gamma_1 \boldsymbol{x}^*_{p_1+1} + \cdots + \gamma_{p_2} \boldsymbol{x}^*_p$. 两个向量之间夹角的余弦正是 $u_\alpha = \boldsymbol{\alpha}' \boldsymbol{x}^{(1)}_\alpha$ 和 $v_\alpha = \boldsymbol{\gamma}' \boldsymbol{x}^{(2)}_\alpha (\alpha = 1, \cdots, N)$ 之间的相关. 寻找 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 使相关极大化等价于在 p_1 空间和 p_2 空间中寻找向量, 它们之间的夹角达到最小 (比如, 余弦是最大). 这给出了第一组典型变量, 第一组典型相关是夹角的余弦. 类似地, 第二组典型变量相应于同第一组典型变量正交的向量, 有最小的夹角.

12.3.2　计算

我们主要用总体数量的方法来讨论计算问题. 可以利用 12.2 节的 (50), (51), (52). $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 的计算可以通过解得 $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{22}\boldsymbol{F}$ 的解 $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ 再乘 $\boldsymbol{\Sigma}_{12}$ 来完成. 如果 p_1 足够小, 行列式 $|\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \nu\boldsymbol{\Sigma}_{11}|$ 可以展开成 ν 的多项式, 可以解关于 ν 的多项式方程. 把解带入 (51) 式来得到向量 $\boldsymbol{\alpha}$.

很多情况下, p_1 太大, 使得这个方法没有效率. 可以用迭代方法

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}(i) = \lambda^2(i+1)\boldsymbol{\alpha}(i+1), \tag{17}$$

从初始的近似 $\boldsymbol{\alpha}(0)$ 开始, 向量 $\boldsymbol{\alpha}(i+1)$ 可以由下式正规化

$$\boldsymbol{\alpha}(i+1)'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}(i+1) = 1. \tag{18}$$

$\lambda^2(i+1)$ 收敛到 λ_1^2, $\boldsymbol{\alpha}(i+1)$ 收敛到 $\boldsymbol{\alpha}^{(1)}$(如果 $\lambda_1 > \lambda_2$), 这可以用主成分所用的方法类似地证明, 用

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{A}\boldsymbol{\Lambda}^2\boldsymbol{A}^{-1}, \tag{19}$$

其来自 12.2 节的 (45) 式. 见习题 12.9.

(19) 式的右边是 $\sum_{i=1}^{p_1} \boldsymbol{\alpha}^{(i)}\lambda_i^2\tilde{\boldsymbol{\alpha}}^{(i)'}$, 其中 $\tilde{\boldsymbol{\alpha}}^{(i)'}$ 是 $\boldsymbol{A}^{-1}$ 的第 i 行. 从 $\boldsymbol{A}'\boldsymbol{\Sigma}_{11}\boldsymbol{A} = \boldsymbol{I}$, 我们得到 $\boldsymbol{A}'\boldsymbol{\Sigma}_{11} = \boldsymbol{A}^{-1}$, 因此 $\boldsymbol{\alpha}^{(i)'}\boldsymbol{\Sigma}_{11} = \tilde{\boldsymbol{\alpha}}^{(i)'}$. 现在有

$$\begin{aligned}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} - \lambda_1^2\boldsymbol{\alpha}^{(1)}\tilde{\boldsymbol{\alpha}}^{(1)'} &= \sum_{i=2}^{p_1} \boldsymbol{\alpha}^{(i)}\lambda_i^2\tilde{\boldsymbol{\alpha}}^{(i)'} \\ &= \boldsymbol{A}\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_{p_1}^2 \end{pmatrix}\boldsymbol{A}^{-1}.\end{aligned} \tag{20}$$

这个矩阵的最大特征根是 λ_2^2. 如果我们现在用这个矩阵迭代, 那么将得到 λ_2^2 和 $\boldsymbol{\alpha}^{(2)}$. 继续这样的方法直到得到足够多的 λ_i^2 和 $\boldsymbol{\alpha}^{(i)}$.

给定 λ_i 和 $\boldsymbol{\alpha}^{(i)}$, 我们就从 $\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}^{(i)} = \lambda_i\boldsymbol{\Sigma}_{22}\boldsymbol{\gamma}^{(i)}$ 或者等价地 $(1/\lambda_i)\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}^{(i)} = \boldsymbol{\gamma}^{(i)}$ 得到 $\boldsymbol{\gamma}^{(i)}$. 可由比较 $\boldsymbol{\Sigma}_{12}\boldsymbol{\gamma}^{(i)}$ 和 $\lambda_i\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}^{(i)}$ 来检验计算.

对样本我们用 $\hat{\boldsymbol{\Sigma}}_{ij}$ 或者 $\boldsymbol{S}_{ij}$ 代替 $\boldsymbol{\Sigma}_{ij}$ 进行计算. 在计算中用 $\boldsymbol{R}_{ij}$ 来得到 $\boldsymbol{S}_1\boldsymbol{a}^{(j)}$ 和 $\mathbf{S}_2\boldsymbol{c}^{(j)}$ 常常会很方便 (因为 $-1 < r_{ij} < 1$), 从这些可以计算出 $\boldsymbol{a}^{(j)}$ 和 $\boldsymbol{c}^{(j)}$.

和主成分的概略类似, 现代的计算方法可用于典型相关和典型变量. 令

$$\boldsymbol{Z}_1 = (\boldsymbol{x}_1^{(1)} - \bar{\boldsymbol{x}}^{(1)}, \cdots, \boldsymbol{x}_N^{(1)} - \bar{\boldsymbol{x}}^{(1)}), \tag{21}$$

$$\boldsymbol{Z}_2 = (\boldsymbol{x}_1^{(2)} - \bar{\boldsymbol{x}}^{(2)}, \cdots, \boldsymbol{x}_N^{(2)} - \bar{\boldsymbol{x}}^{(2)}). \tag{22}$$

这些矩阵转置的 QR 分解 (11.4 节) 是 $\boldsymbol{Z}_i' = \boldsymbol{Q}_i\boldsymbol{R}_i$, 其中 $\boldsymbol{Q}_i'\boldsymbol{Q}_i = \boldsymbol{I}_{p_i}$, $\boldsymbol{R}_i$ 是上三角的. 那么 $\boldsymbol{S}_{ij} = \boldsymbol{Z}_i\boldsymbol{Z}_j' = \boldsymbol{R}_i'\boldsymbol{Q}_i'\boldsymbol{Q}_j\boldsymbol{R}_j, i,j=1,2$, $\boldsymbol{S}_{ii} = \boldsymbol{R}_i'\boldsymbol{R}_i, i=1,2$. 典型相关是 $\boldsymbol{Q}_1'\boldsymbol{Q}_2$ 的奇异值, 也是 $(\boldsymbol{Q}_1'\boldsymbol{Q}_2)(\boldsymbol{Q}_1'\boldsymbol{Q}_2)'$ 的特征根的平方根 (由定理 12.2.2). $\boldsymbol{Q}_1'\boldsymbol{Q}_2$ 的奇异值分解是 $\boldsymbol{P}(\boldsymbol{L}\ \boldsymbol{O})\boldsymbol{T}$, 其中 $\boldsymbol{P}$ 和 $\boldsymbol{T}$ 是正交的, $\boldsymbol{L}$ 是对角的. 为了产生分解, 将 Householder 变换应用于 $\boldsymbol{Q}_1'\boldsymbol{Q}_2$ 的左端和右端, 得到上两对角线矩阵, 即元在主对角线和次主对角线上的矩阵. 吉文斯矩阵用来减小这个矩阵到一个对角矩阵所要求的近似度. 更多的细节可以参考 Kennedy and Gentle(1980) 的 7.2 节和 12.2 节, Chambers(1977), Björck and Golub(1973), Golhb and Luk(1976), Golub and Van Loan(1989).

12.4 统计推断

12.4.1 独立性检验和秩检验

在第 9 章中我们考虑了检验原假设 "$\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 是独立的", 它同原假设 $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$ 等价. 由于 $\boldsymbol{A}'\boldsymbol{\Sigma}_{12}\boldsymbol{\Gamma} = (\boldsymbol{\Lambda}\ \ \boldsymbol{0})$, 可以看出假设等价于 $\boldsymbol{\Lambda} = \boldsymbol{0}$, 即 $\rho_1 = \cdots = \rho_{p_1} = 0$. 检验这个原假设的似然比准则是

$$\frac{\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{vmatrix}}{|\boldsymbol{A}_{11}|\cdot|\boldsymbol{A}_{22}|} = \frac{\begin{vmatrix} \hat{\boldsymbol{A}}' & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{\Gamma}}' \end{vmatrix}}{|\hat{\boldsymbol{A}}'|\cdot|\hat{\boldsymbol{\Gamma}}'|} \cdot \frac{\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{vmatrix}}{|\boldsymbol{A}_{11}|\cdot|\boldsymbol{A}_{22}|} \cdot \frac{\begin{vmatrix} \hat{\boldsymbol{A}} & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{\Gamma}} \end{vmatrix}}{|\hat{\boldsymbol{A}}|\cdot|\hat{\boldsymbol{\Gamma}}|}$$

$$= \frac{\begin{vmatrix} \boldsymbol{I} & \hat{\boldsymbol{\Lambda}} & \boldsymbol{0} \\ \hat{\boldsymbol{\Lambda}} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{vmatrix}}{|\boldsymbol{I}|\cdot|\boldsymbol{I}|} = \begin{vmatrix} \boldsymbol{I} & \hat{\boldsymbol{\Lambda}} \\ \hat{\boldsymbol{\Lambda}} & \boldsymbol{I} \end{vmatrix} = |\boldsymbol{I} - \hat{\boldsymbol{\Lambda}}^2| = \prod_{i=1}^{p_1}(1 - r_i{}^2) \tag{1}$$

的 $N/2$ 次幂, 其中 $r_1 = l_1 \geqslant \cdots \geqslant r_{p_1} = l_{p_1} \geqslant 0$ 是 p_1 个可能的非零样本典型相关. 在原假设下, 对于 -2 倍的似然比准则的对数来说, 其 Barlett 修正的极限分布, 也就是,

$$-\left[N - \frac{1}{2}(p+3)\right]\sum_{i=1}^{p_1}\ln(1 - r_i{}^2), \tag{2}$$

是自由度为 p_1p_2 的 χ^2 分布. (见 9.4 节.) 注意到它近似于

$$N\sum_{i=1}^{p_1} r_i{}^2 = N\mathrm{tr}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{21}^{-1}\boldsymbol{A}_{21}, \tag{3}$$

是 N 倍的 Nagao 准则 [9.5 节的 (2)].

如果 $\mathbf{\Sigma}_{12} \neq \mathbf{0}$, 一个有趣的问题是多少个总体的典型相关不同于零, 即解释 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 之间的相关的需要多少组典型变量? 非零典型相关的数目等于 $\mathbf{\Sigma}_{12}$ 的秩. 检验原假设 $\boldsymbol{H}_k : \rho_{k+1} = \cdots = \rho_{p_1} = 0$ 的似然比准则, 即 $\mathbf{\Sigma}_{12}$ 的秩, 不大于 k, 它是 $\prod_{i=k+1}^{p_1}(1-r_i^2)^{\frac{1}{2}N}$ [Fujikoshi(1974)]. 在原假设下

$$-\left[N-\frac{1}{2}(p+3)\right]\sum_{i=k+1}^{p_1}\ln(1-r_i^2) \tag{4}$$

有近似的自由度是 $(p_1-k)(p_2-k)$ 的 χ^2 分布. [Glynn and Muirhead(1978) 提出用 $N-k-\frac{1}{2}(p+3)+\sum_{i=1}^{k}(1/r_i^2)$ 乘 (4) 中的和; 也可见 Lawlry(1959).]

为了确定总体典型相关中零和非零项的数目, 可以检验所有的根均为 0; 如果假设被拒绝, 检验 p_1-1 个最小根均为零; 依此类推. 当然, 这些步骤并不具有统计独立性, 甚至渐近的统计独立性. 另一个可供选择的方案是, 我们可以同上面相对地用一系列检验: 检验 $\rho_{p_1}=0$, 然后 $\rho_{p_1-1}=\rho_{p_1}=0$, 依此类推. 直到某个假设被拒绝或者直到 $\mathbf{\Sigma}_{12}=\mathbf{0}$ 被接受. 另一个方法 (只用在 p_1 特别小的时候) 是检验 $\rho_{p_1}=0$, 然后是 $\rho_{p_1-1}=0$, 如此等等. 在这个方法中可以用 r_j 检验假设 $\rho_j=0$. 有关的渐近分布将在 12.4.2 节讨论.

12.4.2　典型相关的分布

在 $\mathbf{\Sigma}_{12}=\mathbf{0}$ 的情况下典型相关的密度在 13.4 节中给出, 即所有的总体相关均为 0. 在有些总体相关不是零的情况下的密度由 Constantine(1963) 用矩阵论中超几何函数的方法给出.

大样本理论更易处理. 假设前 k 个典型相关是正的, 小于 1, 均不相等, 假设 p_1-k 个相关是 0. 令

$$\begin{aligned} z_i &= \sqrt{N}\frac{r_i^2-\rho_i^2}{2\rho_i(1-\rho_i^2)}, \qquad i=1,\cdots,k, \\ z_i &= Nr_i^2, \qquad i=k+1,\cdots,p_1 \end{aligned} \tag{5}$$

$z_1,\cdots,z_k$ 的极限分布同 $z_{k+1},\cdots,z_{p_1}$ 的极限分布是互相独立的, z_i 有极限分布 $N(0,1), i=1,\cdots,k$, $z_{k+1},\cdots,z_{p_1}$ 的极限分布的密度是

$$\frac{\pi^{\frac{1}{2}(p_1-k)^2}\exp\left(-\frac{1}{2}\sum_{i=k+1}^{p_1}z_i\right)}{2^{\frac{1}{2}(p_1-k)(p_2-k)}\Gamma_{p_1-k}[\frac{1}{2}(p_1-k)]\Gamma_{p_1-k}[\frac{1}{2}(p_2-k)]} \cdot \prod_{i=k+1}^{p_1} z_i^{\frac{1}{2}(p_2-p_1-1)} \prod_{\substack{i,j=k+1\\ i<j}}^{p_1}(z_i-z_j). \tag{6}$$

这是 13.3 节中有 $W(I_{p_1-k},p_2-k)$ 分布的 (p_1-k) 阶矩阵的特征根的密度 (11). 注意到相应于非零总体相关的平方相关的正规化因子是 $\sqrt{N}$, 而对应于零总体相关的因子是 N. 见第 13 章.

大样本中我们将 r_i^2 视为 $N[\rho_i^2,(1/N)4\rho_i^2(1-\rho_i^2)^2]$ 或者将 r_i 视为 $N[\rho_i,(1/N)(1-\rho_i^2)^2]$ (由定理 4.2.3) 来得到 ρ_i 的检验或者 ρ_i 的置信区间. Lawley(1959) 证明了变换 $z_i=\tanh^{-1}(r_i)$[见 4.2.3 节] 并不能使方差稳定并且在估计 $\zeta_i=\tanh^{-1}(\rho_1)$ 时有显著的偏差.

12.5 一个例子

本节我们考虑一个简单的示例. Rao[(1952), p. 245] 给出了 25 个家庭中成年长子和次子的一些测量数据作为样本. (习题 3.1 和习题 4.41 已经用过.) 令第 α 个家庭长子的头长为 $x_{1\alpha}$, 长子的头宽为 $x_{2\alpha}$, 次子的头长为 $x_{3\alpha}$, 头宽为 $x_{4\alpha}$. 我们想研究长子和次子的测量数据之间的关系. 这样 $x_\alpha^{(1)'}=(x_{1\alpha},x_{2\alpha})$, $x_\alpha^{(2)'}=(x_{3\alpha},x_{4\alpha})$. 数据可以概括为 ①

$$\bar{x}'=(185.72,151.12,183.84,149.24),$$

$$\boldsymbol{S}=\begin{pmatrix}95.2933 & 52.8683 & 69.6617 & 46.1117\\ 52.8683 & 54.3600 & 51.3117 & 35.0533\\ 69.6617 & 51.3117 & 100.8067 & 56.5400\\ 46.1117 & 35.0533 & 56.5400 & 45.0233\end{pmatrix}=\begin{bmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{bmatrix}. \tag{1}$$

相关阵是

$$\boldsymbol{R}=\begin{pmatrix}1.0000 & 0.7346 & \vdots & 0.7108 & 0.7040\\ 0.7346 & 1.0000 & \vdots & 0.6932 & 0.7086\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0.7108 & 0.6932 & \vdots & 1.0000 & 0.8392\\ 0.7040 & 0.7086 & \vdots & 0.8392 & 1.0000\end{pmatrix}=\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{pmatrix}. \tag{2}$$

除了次子的两项测量的相关系数外, 其余所有的相关系数都在 0.7 附近. 特别地, $\boldsymbol{R}_{12}$ 接近秩 1, 因此第二组典型相关接近零.

我们计算得

$$\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}=\begin{pmatrix}0.405\,769 & 0.333\,205\\ 0.363\,480 & 0.428\,976\end{pmatrix}, \tag{3}$$

$$\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}=\begin{pmatrix}0.544\,311 & 0.538\,841\\ 0.538\,841 & 0.534\,950\end{pmatrix}. \tag{4}$$

行列式方程是

① Rao 的计算有错误, 他的最后的 "差" 不正确.

$$0=\begin{vmatrix} 0.544\,311-1.0000\nu & 0.538\,841-0.7346\nu \\ 0.538\,841-0.7346\nu & 0.534\,950-1.0000\nu \end{vmatrix}$$
$$=0.460\,363\nu^2-0.287\,596\nu+0.000\,830. \tag{5}$$

根是 0.621 816 和 0.002 900, 因此 $l_1=0.788\,553$, $l_2=0.053\,852$. 相应于这些根的向量是

$$\boldsymbol{S}_1\boldsymbol{a}^{(1)}=\begin{pmatrix} 0.552\,166 \\ 0.521\,548 \end{pmatrix}, \qquad \boldsymbol{S}_1\boldsymbol{a}^{(2)}=\begin{pmatrix} 1.366\,501 \\ -1.378\,467 \end{pmatrix}, \tag{6}$$

其中

$$\boldsymbol{S}_1=\begin{pmatrix} \sqrt{s_{11}} & 0 \\ 0 & \sqrt{s_{22}} \end{pmatrix}=\begin{pmatrix} 9.7618 & 0 \\ 0 & 7.3729 \end{pmatrix}. \tag{7}$$

我们对 $\boldsymbol{S}_1\boldsymbol{a}^{(i)}$ 利用 $(1/l_i)\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}$ 来得到

$$\boldsymbol{S}_2\boldsymbol{c}^{(1)}=\begin{pmatrix} 0.504\,511 \\ 0.538\,242 \end{pmatrix}, \quad \boldsymbol{S}_2\boldsymbol{c}^{(2)}=\begin{pmatrix} 1.767\,281 \\ -1.757\,288 \end{pmatrix}, \tag{8}$$

其中

$$\boldsymbol{S}_2=\begin{pmatrix} \sqrt{s_{33}} & 0 \\ 0 & \sqrt{s_{44}} \end{pmatrix}=\begin{pmatrix} 10.0402 & 0 \\ 0 & 6.7099 \end{pmatrix}. \tag{9}$$

我们通过计算

$$\frac{1}{l_1}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}(\boldsymbol{S}_2\boldsymbol{c}^{(1)})=\begin{pmatrix} 0.552\,157 \\ 0.521\,560 \end{pmatrix}, \quad \frac{1}{l_2}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}(\boldsymbol{S}_2\boldsymbol{c}^{(2)})=\begin{pmatrix} 1.365\,151 \\ -1.376\,741 \end{pmatrix} \tag{10}$$

来检查这些计算. (10) 式的第一个向量接近于对应 (6) 式的第一个向量; 事实上, 它是一个微小的改进, 计算等价于对 $\boldsymbol{S}_1\boldsymbol{a}^{(1)}$ 的一个迭代. (10) 式的第二个向量接近于对应 (6) 的第二个向量. 一个原因是 l_2 仅对四五个有效的位数才正确 (如 $\nu_2=l_2^2$), 因此 $\boldsymbol{S}_2\boldsymbol{c}^{(2)}$ 的元也仅仅对同样有效的位数是正确的; 第二, $\boldsymbol{S}_2\boldsymbol{c}^{(2)}$ 对应于较小的根的事实意味着迭代减少了精度而不是增大了精度. 我们最后的结果是

$$\begin{array}{ccc} & (1) & (2) \\ l_i= & 0.789, & 0.054, \\ \boldsymbol{a}^{(i)}= & \begin{pmatrix} 0.0566 \\ 0.0707 \end{pmatrix}, & \begin{pmatrix} 0.1400 \\ -0.1870 \end{pmatrix}, \\ \boldsymbol{c}^{(i)}= & \begin{pmatrix} 0.0502 \\ 0.0802 \end{pmatrix}, & \begin{pmatrix} 0.1760 \\ -0.2619 \end{pmatrix}. \end{array} \tag{11}$$

两个典型相关中较大的, 0.789, 比一组变量同另一组变量之间个体的相关要大. 第二个典型相关非常接近零. 这意味着, 为了研究长子和次子的两个测量的关系, 我们可以把注意力放在第一组典型变量上; 第二组典型变量只有微小的相关. 第一组典型变量与两组测量数据除以各自的标准差之后的和是近似成比例的; 第二组典型变量同两个标准度量的差近似成比例.

12.6 线性相关期望值

12.6.1 回归矩阵的典型分析

本节中我们讨论一个随机向量和一个非随机向量的典型相关和典型变量. 随机向量的期望值是非随机向量的一个线性函数 (第 8 章). 我们找到了新的坐标系, 在此坐标系下随机向量的每个坐标的期望值只取决于非随机向量的一个坐标; 随机向量的坐标在随机意义下是不相关的, 非随机向量的坐标在样本中是不相关的. 坐标的次序依据相对于方差的平方和效应. 代数方法类似于 12.2 节中介绍的.

如果 $\boldsymbol{X}$ 有正态分布 $N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\mathbf{X},\boldsymbol{\mu},\boldsymbol{\Sigma}$ 有如 12.2 节 (1) 和 (2) 式的划分, $\boldsymbol{\mu}=(\boldsymbol{\mu}^{(1)\prime},\boldsymbol{\mu}^{(2)\prime})'$, 给定 $\boldsymbol{x}^{(2)}$ 下 $\boldsymbol{X}^{(1)}$ 的条件分布是正态的, 有均值

$$\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}),\qquad \boldsymbol{\beta}=\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1},\tag{1}$$

协方差阵

$$\boldsymbol{\Sigma}_{11.2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.\tag{2}$$

我们考虑一组随机向量 $\boldsymbol{X}_1^{(1)},\cdots,\boldsymbol{X}_N^{(1)}$, 它们有取决于 $\boldsymbol{x}_1^{(2)},\cdots,\boldsymbol{x}_N^{(2)}$ 的期望值 (非随机的), 我们可将 $\boldsymbol{X}_\phi^{(1)}$ 的条件期望值写为 $\boldsymbol{\tau}+\boldsymbol{\beta}(\boldsymbol{x}_\phi^{(2)}-\bar{\boldsymbol{x}}^{(2)})$, 其中 $\boldsymbol{\tau}=\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\bar{\boldsymbol{x}}^{(2)}-\boldsymbol{\mu}^{(2)})$ 可以当作一个参数向量. 这是 8.2 节的模型, 只是在记法上有一个微小的变化.

本节的模型是

$$\mathrm{E}(\boldsymbol{X}_\phi^{(1)})=\boldsymbol{\tau}+\boldsymbol{\beta}(\boldsymbol{x}_\phi^{(2)}-\bar{\boldsymbol{x}}^{(2)}),\qquad \phi=1,\cdots,N,\tag{3}$$

其中 $\boldsymbol{x}_1^{(2)},\cdots,\boldsymbol{x}_N^{(2)}$ 是一组非随机向量 $(q\times 1)$, $\bar{\boldsymbol{x}}^{(2)}=N^{-1}\sum_{\phi=1}^N\boldsymbol{x}_\phi^{(2)}$. 协方差阵是

$$\mathrm{E}(\boldsymbol{X}_\phi^{(1)}-\mathrm{E}(\boldsymbol{X}_\phi^{(1)}))(\boldsymbol{X}_\phi^{(1)}-\mathrm{E}(\boldsymbol{X}_\phi^{(1)}))'=\boldsymbol{\Psi}.\tag{4}$$

考虑 $\boldsymbol{X}_\phi^{(1)}$ 的一组线性组合, 令 $\boldsymbol{U}_\phi=\boldsymbol{\alpha}'\boldsymbol{X}_\phi^{(1)}$. 则 $\boldsymbol{U}_\phi$ 的方差是 $\boldsymbol{\alpha}'\boldsymbol{\Psi}\boldsymbol{\alpha}$, 期望值是

$$\mathrm{E}(\boldsymbol{U}_\phi)=\boldsymbol{\alpha}'\boldsymbol{\tau}+\boldsymbol{\alpha}'\boldsymbol{\beta}(\boldsymbol{x}_\phi^{(2)}-\bar{\boldsymbol{x}}^{(2)}).\tag{5}$$

平均期望值是 $(1/N)\sum_{\phi=1}^N\mathrm{E}(\boldsymbol{U}_\phi)=\boldsymbol{\alpha}'\boldsymbol{\tau}$, 由 $\boldsymbol{x}^{(2)}$ 导致的平均平方和是

$$\begin{aligned}\frac{1}{n}\sum_{\phi=1}^N(\mathrm{E}(\boldsymbol{U}_\phi)-\boldsymbol{\alpha}'\boldsymbol{\tau})^2&=\frac{1}{n}\sum_{\phi=1}^N\boldsymbol{\alpha}'\boldsymbol{\beta}(x_\phi^{(2)}-\bar{x}^{(2)})(x_\phi^{(2)}-\bar{x}^{(2)})'\boldsymbol{\beta}'\boldsymbol{\alpha}\\&=\boldsymbol{\alpha}'\boldsymbol{\beta}S_{22}\boldsymbol{\beta}'\boldsymbol{\alpha}.\end{aligned}\tag{6}$$

我们寻找使平均平方和相对于它的方差达到极大的线性组合, 即相依变量的线性组合对独立变量有最大化效应. 在满足 $\boldsymbol{\alpha}'\boldsymbol{\Psi}\boldsymbol{\alpha}=1$ 的条件下要使 (6) 式极大化. 这就得到了向量方程

$$(\boldsymbol{\beta}S_{22}\boldsymbol{\beta}'-\kappa\boldsymbol{\Psi})\boldsymbol{\alpha}=\mathbf{0},\tag{7}$$

其中 κ 满足

$$|\boldsymbol{\beta}S_{22}\boldsymbol{\beta}'-\kappa\boldsymbol{\Psi}|=0. \tag{8}$$

(7) 式左乘 $\boldsymbol{\alpha}'$ 得到 $\boldsymbol{\alpha}'\boldsymbol{\beta}S_{22}\boldsymbol{\beta}'\boldsymbol{\alpha}=\kappa$, $\boldsymbol{\alpha}$ 和 κ 满足 $\boldsymbol{\alpha}'\boldsymbol{\Psi}\boldsymbol{\alpha}=1$ 和 (7) 式; 为了求得极大值, 我们取 (8) 式的最大根, 设为 κ_1. 用 $\boldsymbol{\alpha}^{(1)}$ 表示这个向量, 则相应的随机变量为 $\boldsymbol{U}_{1\phi}=\boldsymbol{\alpha}^{(1)'}\mathbf{X}_{\phi}^{(1)}$. 第一个典型变量的期望值是 $\mathrm{E}(\boldsymbol{U}_{1\phi})=\boldsymbol{\alpha}^{(1)'}[\boldsymbol{\beta}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})+\tau]$. 令 $\boldsymbol{\alpha}^{(1)'}\boldsymbol{\beta}=k\boldsymbol{\gamma}^{(1)'}$, 其中 k 由

$$\begin{aligned}1&=\frac{1}{n}\sum_{\phi=1}^{N}\left(\boldsymbol{\gamma}^{(1)'}\boldsymbol{x}_{\phi}^{(2)}-\frac{1}{N}\sum_{\eta=1}^{N}\boldsymbol{\gamma}^{(1)'}\boldsymbol{x}_{\eta}^{(2)}\right)^2 \\ &=\frac{1}{n}\sum_{\phi=1}^{N}\boldsymbol{\gamma}^{(1)'}\left(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)}\right)\left(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)}\right)'\boldsymbol{\gamma}^{(1)} \\ &=\boldsymbol{\gamma}^{(1)'}S_{22}\boldsymbol{\gamma}^{(1)}\end{aligned} \tag{9}$$

确定. 那么 $k=\sqrt{\kappa_1}$. 令 $\boldsymbol{U}_{1\phi}=\boldsymbol{\gamma}^{(1)'}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$. 那么 $\mathrm{E}(\boldsymbol{U}_{1\phi})=\sqrt{\kappa_1}v_{\phi}^{(1)}+\boldsymbol{\alpha}^{(1)'}\boldsymbol{\tau}$.

接下来设我们已经得到了一个线性组合 $\boldsymbol{U}_{\phi}=\boldsymbol{\alpha}'\boldsymbol{X}_{\phi}^{(1)}$, 它在所有方差是 1 并且同 $\boldsymbol{U}_{1\phi}$ 不相关的所有线性组合中, 有最大的平方和效应, 即 $0=\mathrm{E}(\boldsymbol{U}_{\phi}-\mathrm{E}(\boldsymbol{U}_{\phi}))(\boldsymbol{U}_{1\phi}-\mathrm{E}(\boldsymbol{U}_{1\phi}))'=\boldsymbol{\alpha}'\boldsymbol{\Psi}\boldsymbol{\alpha}^{(1)}$. 如 12.2 节, 我们可以用拉格朗日乘数建立这个极大化问题, 对某些满足 (8) 式和 $\boldsymbol{\alpha}'\boldsymbol{\Psi}\boldsymbol{\alpha}=1$ 的 κ, 找到 $\boldsymbol{\alpha}$ 满足 (7) 式. 用同 12.2 节类似的方式继续这个方法. 我们概括一下结果.

第 j 个典型随机变量是 $\boldsymbol{U}_{j\phi}=\boldsymbol{\alpha}^{(j)'}\boldsymbol{X}_{\phi}^{(1)}$, 其中 $\boldsymbol{\alpha}^{(j)}$ 对 $\kappa=\kappa_j$ 满足 (7) 式和 $\boldsymbol{\alpha}^{(j)'}\boldsymbol{\Psi}\boldsymbol{\alpha}^{(j)}=1$; $\kappa_1\geqslant\kappa_2\geqslant\cdots\geqslant\kappa_{p_1}$ 是 (8) 式的根. 我们假定 $\boldsymbol{\beta}$ 的秩 $p_1\leqslant p_2$. (则 $\kappa_{p_1}>0$.)$\boldsymbol{U}_{j\phi}$ 在所有有单位方差并且与 $\boldsymbol{U}_{1\phi},\cdots,\boldsymbol{U}_{j-1,\phi}$ 不相关的线性组合中有最大的平方和效应. 令 $\boldsymbol{\gamma}^{(j)}=(1/\sqrt{\kappa_j})\boldsymbol{\beta}'\boldsymbol{\alpha}^{(j)}, v_j=\boldsymbol{\alpha}^{(j)'}\boldsymbol{\tau}, v_{j\phi}=\boldsymbol{\gamma}^{(j)'}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$. 那么

$$\mathrm{E}(\boldsymbol{U}_{j\phi})=\sqrt{\kappa_j}v_{j\phi}+v_j, \tag{10}$$

$$\frac{1}{n}\sum_{\phi=1}^{N}\left(v_{j\phi}-\frac{1}{N}\sum_{\eta=1}^{N}v_{i\eta}\right)^2=1, \tag{11}$$

$$\sum_{\phi=1}^{N}\left(v_{j\phi}-\frac{1}{N}\sum_{\eta=1}^{N}v_{j\eta}\right)\left(v_{j\phi}-\frac{1}{N}\sum_{\eta=1}^{N}v_{j\eta}\right)=0,\quad i\neq j. \tag{12}$$

如果 $p_2>p_1$, 那么可以选择 $\boldsymbol{\gamma}^{(p_1+1)},\cdots,\boldsymbol{\gamma}^{(p_2)}$ 使得 $v_{p_1+1,\phi}=\boldsymbol{\gamma}^{(p_1+1)'}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)}),\cdots,v_{p_2\phi}=\boldsymbol{\gamma}^{(p_2)'}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$ 满足 (11) 式和 (12) 式.

令 $\boldsymbol{A}=(\boldsymbol{\alpha}^{(1)},\cdots,\boldsymbol{\alpha}^{(p_1)}),\boldsymbol{\Gamma}_1=(\boldsymbol{\gamma}^{(1)},\cdots,\boldsymbol{\gamma}^{(p_1)}),\boldsymbol{\Gamma}_2=(\boldsymbol{\gamma}^{(p_1+1)},\cdots,\boldsymbol{\gamma}^{(p_2)}),\boldsymbol{\Delta}=\mathrm{diag}(\delta_1,\cdots,\delta_{p_1})=\mathrm{diag}(\sqrt{\kappa_1},\cdots,\sqrt{\kappa_{p_1}}),\boldsymbol{U}_{\phi}=\boldsymbol{A}'\boldsymbol{X}_{\phi}^{(1)},\boldsymbol{v}_{\phi}^{(1)}=\boldsymbol{\Gamma}_1'(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)}),\boldsymbol{v}_{\phi}^{(2)}=\boldsymbol{\Gamma}_2(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})$, $\boldsymbol{v}_{\phi}=(\boldsymbol{v}_{\phi}^{(1)'},\ \boldsymbol{v}_{\phi}^{(2)'})',\phi=1,\cdots,N$. 那么有

$$\mathrm{E}(\boldsymbol{U}_{\phi}-\mathrm{E}(\boldsymbol{U}_{\phi}))(\boldsymbol{U}_{\phi}-\mathrm{E}(\boldsymbol{U}_{\phi}))'=\mathbf{A}'\boldsymbol{\Psi}\mathbf{A}=\boldsymbol{I}. \tag{13}$$

$$\mathrm{E}\boldsymbol{U}_\phi = \boldsymbol{\Delta v}_\phi^{(1)} + v, \quad \phi = 1, \cdots, N, \tag{14}$$

$$\frac{1}{n}\sum_{\phi=1}^{N}\left(v_\phi - \frac{1}{N}\sum_{\eta=1}^{N} v_\eta\right)\left(v_\phi - \frac{1}{N}\sum_{\eta=1}^{N} v_\eta\right)' = \boldsymbol{I} \tag{15}$$

随机典型变量有方差 1 并且不相关. 每个随机典型变量的期望值是相应的非随机典型变量乘一个常数的倍数. 非随机典型变量有样本方差 1 并且在样本中不相关.

如果 $p_1 > p_2$, $\boldsymbol{\beta}$ 的极大秩为 p_2, $\kappa_{p_2+1} = \cdots = \kappa_{p_1} = 0$. 这种情况下, 我们定义 $\boldsymbol{A}_1 = (\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(p_2)}), \boldsymbol{A}_2 = (\boldsymbol{\alpha}^{(p_2+1)}, \cdots, \boldsymbol{\alpha}^{(p_1)})$, 其中 $\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(p_2)}$(对应于正的 κ) 如前面所定义, $\boldsymbol{\alpha}^{(p_2+1)}, \cdots, \boldsymbol{\alpha}^{(p_1)}$ 是任意满足 $\boldsymbol{\alpha}^{(j)\prime}\boldsymbol{\Psi}\boldsymbol{\alpha}^{(j)} = 1$ 和 $\boldsymbol{\alpha}^{(j)\prime}\boldsymbol{\Psi}\boldsymbol{\alpha}^{(i)} = 0 (i \neq j)$ 的向量. 那么 $\mathrm{E}(\boldsymbol{U}_\phi^{(i)}) = \delta_i \boldsymbol{v}_\phi^{(i)} + \boldsymbol{v}_i, i = 1, \cdots, p_2$, $\mathrm{E}(\boldsymbol{U}_\phi^{(i)}) = \boldsymbol{v}_i, i = p_2 + 1, \cdots, p_1$.

不管哪种情况, 如果 $\boldsymbol{\beta}$ 的秩 $r < \min(p_1, p_2)$, 则 (8) 式有 r 个非零的根, 因此 $\mathrm{E}(\boldsymbol{U}_\phi^{(i)}) = \delta_i \boldsymbol{v}_\phi^{(i)} + \boldsymbol{v}_i$, $i = 1, \cdots, r$.

12.6.2 估计

令 $\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_N^{(1)}$ 是有 12.6.1 节中提到的概率结构的 $\boldsymbol{X}_1^{(1)}, \cdots, \boldsymbol{X}_N^{(1)}$ 的一组观测, 令 $\boldsymbol{x}_1^{(2)}, \cdots, \boldsymbol{x}_N^{(2)}$ 是相应的一组独立变量. 我们可以用

$$\hat{\boldsymbol{\tau}} = \frac{1}{N}\sum_{\phi=1}^{N} \boldsymbol{x}_\phi^{(1)} = \bar{\boldsymbol{x}}^{(1)}, \tag{16}$$

$$\hat{\boldsymbol{\beta}} = \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} = \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}, \tag{17}$$

$$\begin{aligned}\tilde{\boldsymbol{\Psi}} &= \frac{1}{n}\sum_{\phi=1}^{N}\left[\boldsymbol{x}_\phi^{(1)} - \bar{\boldsymbol{x}}^{(1)} - \hat{\boldsymbol{\beta}}(\boldsymbol{x}_\phi^{(2)} - \bar{\boldsymbol{x}}^{(2)})\right]\left[\boldsymbol{x}_\phi^{(1)} - \bar{\boldsymbol{x}}^{(1)} - \hat{\boldsymbol{\beta}}(\boldsymbol{x}_\phi^{(2)} - \bar{\boldsymbol{x}}^{(2)})\right]' \\ &= \frac{1}{n}(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}) = \boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\end{aligned} \tag{18}$$

来估计 $\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\Psi}$, 其中所有的 $\boldsymbol{A}$ 和 $\boldsymbol{S}$ 由前面定义. (用 $n = N - 1$ 除比用 N 除更合适, 后面将会导出极大似然估计量.)

(7) 式和 (8) 式的样本模拟是

$$\begin{aligned}\boldsymbol{0} &= (\hat{\boldsymbol{\beta}}\boldsymbol{S}_{22}\hat{\boldsymbol{\beta}}' - k\tilde{\boldsymbol{\Psi}})\bar{\boldsymbol{a}} \\ &= [\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} - k(\boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21})]\bar{\boldsymbol{a}},\end{aligned} \tag{19}$$

$$\begin{aligned}0 &= |\hat{\boldsymbol{\beta}}\boldsymbol{S}_{22}\hat{\boldsymbol{\beta}}' - k\tilde{\boldsymbol{\Psi}}| \\ &= |\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21} - k(\boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21})|.\end{aligned} \tag{20}$$

(20) 的根 $k_1 \geqslant \cdots \geqslant k_{p_1}$ 估计了 (8) 的根 $\kappa_1 \geqslant \cdots \geqslant \kappa_{p_1}$, 对应的 (19) 的解 $\bar{\boldsymbol{a}}^{(1)}, \cdots, \bar{\boldsymbol{a}}^{(p_1)}$, 由 $\bar{\boldsymbol{a}}^{(i)\prime}\tilde{\boldsymbol{\Psi}}\bar{\boldsymbol{a}}^{(i)} = 1$ 正规化, 估计 $\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(p_1)}$. 那么 $\bar{\boldsymbol{c}}^{(j)} = (1/\sqrt{k_j})\hat{\boldsymbol{\beta}}'\bar{\boldsymbol{a}}^{(j)}$ 估计 γ_j, $n_j = \bar{\boldsymbol{a}}^{(j)\prime}\bar{\boldsymbol{x}}^{(1)}$ 估计 v_j. 样本典型变量是 $\bar{\boldsymbol{a}}^{(j)\prime}\boldsymbol{x}_\phi{}^{(1)}$ 和 $\bar{\boldsymbol{c}}^{(j)\prime}(\boldsymbol{x}_\phi{}^{(2)} - \bar{\boldsymbol{x}}^{(2)}), j = 1, \cdots, p_1, \phi = 1, \cdots, N$. 如果 $p_1 > p_2$, 那么另外 $p_1 - p_2$ 个 $\bar{\boldsymbol{a}}^{(j)\prime}$ 可以定义为满足 $\bar{\boldsymbol{a}}^{(j)\prime}\tilde{\boldsymbol{\Psi}}\bar{\boldsymbol{a}}^{(j)} = 1$ 和 $\bar{\boldsymbol{a}}^{(j)\prime}\tilde{\boldsymbol{\Psi}}\bar{\boldsymbol{a}}^{(i)} = 0, i \neq j$.

12.6.3 典型变量间的关系

在 12.3 节, 定义根 $l_1 \geqslant \cdots \geqslant l_{p_1}$ 满足

$$0=\begin{vmatrix} -l\boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & -l\boldsymbol{S}_{22} \end{vmatrix}=(-1)^{p_2}l^{p_2-p_1}|\boldsymbol{S}_{22}||\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}-l^2\boldsymbol{S}_{11}|. \tag{21}$$

因为 (20) 式可以写成

$$0=|(1+k)\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}-k\boldsymbol{S}_{11}|, \tag{22}$$

我们可以得到 $l_i^2=k_i/(1+k_i)$ 和 $k_i=l_i^2/(1-l_i^2), i=1,\cdots,p_1$. 12.3 节中的向量 $\boldsymbol{a}^{(i)}$ 满足

$$\begin{aligned}\mathbf{0}&=\left(\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}-l_i^2\boldsymbol{S}_{11}\right)\boldsymbol{a}^{(i)} \\ &=\left(\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}-\frac{k_i}{1+k_i}\boldsymbol{S}_{11}\right)\boldsymbol{a}^{(i)} \\ &=\frac{1}{1+k_i}\left[\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}-k_i(\boldsymbol{S}_{11}-\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21})\right]\boldsymbol{a}^{(i)},\end{aligned} \tag{23}$$

当 $k=k_i$ 时 (23) 式和 (19) 式等价. 同正规化的 $\boldsymbol{a}^{(i)\prime}\boldsymbol{S}_{11}\boldsymbol{a}^{(i)}=1$ 和 $\bar{\boldsymbol{a}}^{(i)\prime}(\boldsymbol{S}_{11}-\boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21})\bar{\boldsymbol{a}}^{(i)}=1$ 比较, 就得到 $\bar{\boldsymbol{a}}^{(i)}=(1/\sqrt{1-l_i^2})\boldsymbol{a}^{(i)}$. 那么 $\bar{\boldsymbol{c}}^{(j)}=(1/\sqrt{k_j})\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\bar{\boldsymbol{a}}^{(j)}=\boldsymbol{c}^{(j)}$.

我们看到, 典型变量分析可以应用到两个向量有联合随机分布的情况, 或者一个向量是随机的, 另一个向量是非随机的情况. 两种方法定义的典型变量除了正规化之外是相同的. 相应的典型变量之间的关系的度量也就是 (典型) 关系或者是 "解释方差" 对 "非解释方差" 的比.

12.6.4 秩检验

非零根 κ_j 的个数是回归矩阵 $\boldsymbol{\beta}$ 的秩, 它是需要用来表示 $\boldsymbol{X}_\phi^{(1)}$ 的期望值的回归变量的线性组合的个数. 我们可以检验假设秩是 k(如果 $p_1 \leqslant p_2, 1\leqslant k\leqslant p_1$) 和备择假设秩 $\geqslant k$. 原假设是

$$H_k:\kappa_{k+1}=\cdots=\kappa_{p_1}=0. \tag{24}$$

似然比准则 [Anderson(1951b)] 是

$$\prod_{i=k+1}^{p_1}(1+k_i)^{-1}=\prod_{i=k+1}^{p_1}(1+l_i^2) \tag{25}$$

的幂. 注意到这个准则与两个向量都是随机的情况相同 (12.4 节). 则

$$-\left[N-\frac{1}{2}(p+3)\right]\sum_{i=k+1}^{p_1}\ln(1-l_i^2) \tag{26}$$

有近似的自由度是 $(p_1-k)(p_2-k)$ 的 χ^2 分布.

在 0 和 p_1 之间的秩的确定可以用 12.4 节的方法.

12.6.5 线性函数关系

12.6 节的研究可以用其他的方法实施. 举例来说, 可以建立平衡单向方差分析

$$\boldsymbol{Y}_{\alpha j}=\boldsymbol{\nu}_{\alpha}+\boldsymbol{\mu}+\boldsymbol{U}_{\alpha j},\qquad \alpha=1,\cdots,m,\ j=1,\cdots,l, \tag{27}$$

其中 $\mathrm{E}(\boldsymbol{U}_{\alpha})=\boldsymbol{0},\mathrm{E}(\boldsymbol{U}_{\alpha}\boldsymbol{U}_{\alpha}')=\boldsymbol{\Psi},\sum_{\alpha=1}^{m}\boldsymbol{\nu}_{\alpha}=\boldsymbol{0}$,

$$\boldsymbol{\Theta\nu}_{\alpha}=\boldsymbol{0},\qquad \alpha=1,\cdots,m, \tag{28}$$

其中 $\boldsymbol{\Theta}$ 是 $q\times p_1$ 矩阵且秩为 $q(<p_1)$. 这是 12.6.1 节的模型在 $\phi=1,\cdots,N$ 时的一种特殊情况, 被一组指标 (α,j), $\boldsymbol{X}_{\phi}^{(1)}=\boldsymbol{Y}_{\alpha j},\boldsymbol{\tau}=\boldsymbol{\mu},\boldsymbol{\beta}(\boldsymbol{x}_{\phi}^{(2)}-\bar{\boldsymbol{x}}^{(2)})=\boldsymbol{\nu}_{\alpha}$ 用 8.8 节中的哑变量所取代. $(\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_m)$ 的秩是 $\boldsymbol{\beta}$ 的秩, 即 $r=p_1-q$. (8) 式有 q 个等于零的根,

$$\boldsymbol{\beta}\boldsymbol{S}_{22}\boldsymbol{\beta}'=l\sum_{\alpha=1}^{m}\boldsymbol{\nu}_{\alpha}\boldsymbol{\nu}_{\alpha}'. \tag{29}$$

模型 (27) 可以解释为对 $\boldsymbol{\nu}_{\alpha}+\boldsymbol{\mu}$ 含误差项的重复观测. (28) 式的分量方程是线性函数关系.

令 $\bar{\boldsymbol{y}}_{\alpha}=(1/l)\sum_{j=1}^{l}\boldsymbol{y}_{\alpha j}$, $\bar{\boldsymbol{y}}=(1/m)\sum_{\alpha=1}^{m}\bar{\boldsymbol{y}}_{\alpha}$. 效应的平方和是

$$\boldsymbol{H}=l\sum_{\alpha=1}^{m}(\bar{\boldsymbol{y}}_{\alpha}-\bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_{\alpha}-\bar{\boldsymbol{y}})'=n\hat{\boldsymbol{\beta}}\boldsymbol{S}_{22}\hat{\boldsymbol{\beta}}', \tag{30}$$

有自由度 $m-1$, 误差的平方和是

$$\boldsymbol{G}=\sum_{\alpha=1}^{m}\sum_{j=1}^{l}(\boldsymbol{y}_{\alpha j}-\bar{\boldsymbol{y}}_{\alpha})(\boldsymbol{y}_{\alpha j}-\bar{\boldsymbol{y}}_{\alpha})'=n\tilde{\boldsymbol{\Psi}}, \tag{31}$$

自由度是 $m(l-1)$. $p_1<P_2$ 的情况对应于 $p_1<l$ 的情况. 则 $\boldsymbol{\Theta}$ 的一个极大似然估计量是

$$\hat{\boldsymbol{\Theta}}=(\bar{\boldsymbol{a}}^{(r+1)},\cdots,\bar{\boldsymbol{a}}^{(p_1)})', \tag{32}$$

$\boldsymbol{\nu}_{\alpha}$ 的极大似然估计量是

$$\hat{\boldsymbol{\nu}}_{\alpha}=\tilde{\boldsymbol{\Psi}}\hat{\boldsymbol{\Theta}}'\hat{\boldsymbol{\Theta}}(\bar{\boldsymbol{y}}_{\alpha}-\bar{\boldsymbol{y}}),\quad \alpha=1,\cdots,n. \tag{33}$$

估计量 (32) 可以左乘任意非奇异 $q\times q$ 矩阵从而得到另一个估计量. 更完全的讨论, 见 Anderson(1984a) 和 Kendall and Stuart(1973).

12.7 降秩回归

降秩回归包括在 $\mathrm{E}(\boldsymbol{X}^{(1)}|\boldsymbol{X}^{(2)})=\boldsymbol{\beta}\boldsymbol{X}^{(2)}$ 中用秩为 k 的矩阵 $\hat{\boldsymbol{\beta}}$ 估计回归矩阵 $\boldsymbol{\beta}$. 在估计一个联立方程系统的一部分 (单个方程) 的极限信息极大似然方法中 (12.8 节), 假设回归矩阵的秩比矩阵的阶少一. Anderson(1951a) 获得了 $\boldsymbol{\beta}$ 的极大似然估计量, 其模型为

$$\boldsymbol{X}_{\alpha}^{(1)}=\boldsymbol{\tau}+\boldsymbol{\beta}(\boldsymbol{x}_{\alpha}^{(2)}-\bar{\boldsymbol{x}}^{(2)})+\boldsymbol{Z}_{\alpha},\qquad \alpha=1,\cdots,N, \tag{1}$$

指定 $\boldsymbol{\beta}$ 的秩是 k ($\leqslant p_1$), 向量 $\boldsymbol{x}_1^{(2)},\cdots,\boldsymbol{x}_N^{(2)}$ 是非随机的, $\boldsymbol{Z}_\alpha$ 是正态分布的. 基于样本 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$, $\hat{\boldsymbol{\Sigma}}$ 如 12.3 节 (2) 式定义, $\hat{\boldsymbol{\Lambda}},\hat{\boldsymbol{A}},\hat{\boldsymbol{\Gamma}}$ 如 (3), (4), (5) 定义. 分成 $\hat{\boldsymbol{\Lambda}}=\mathrm{diag}(\hat{\boldsymbol{\Lambda}}_1,\hat{\boldsymbol{\Lambda}}_2)$, $\hat{\boldsymbol{A}}=(\hat{\boldsymbol{A}}_1,\hat{\boldsymbol{A}}_2),\hat{\boldsymbol{\Gamma}}=(\hat{\boldsymbol{\Gamma}}_1,\hat{\boldsymbol{\Gamma}}_2)$, 其中 $\hat{\boldsymbol{\Lambda}}_1,\hat{\boldsymbol{A}}_1,\hat{\boldsymbol{\Gamma}}_1$ 有 k 列. 令 $\hat{\boldsymbol{\Phi}}_1=\hat{\boldsymbol{A}}_1(\boldsymbol{I}_k-\hat{\boldsymbol{\Lambda}}_1^2)^{-\frac{1}{2}}$.

定义 12.7.1 (降秩回归) (1) 式的降秩回归估计量是

$$\hat{\boldsymbol{B}}_k=\boldsymbol{S}_{YX}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}_1'=\boldsymbol{S}_{YY}\hat{\boldsymbol{A}}_1\hat{\boldsymbol{\Lambda}}_1\hat{\boldsymbol{\Gamma}}_1'=\hat{\boldsymbol{\Sigma}}_{\hat{z}\hat{z}}\hat{\boldsymbol{\Phi}}_1\hat{\boldsymbol{\Phi}}_1\boldsymbol{B}. \tag{2}$$

其中$\boldsymbol{B}=\hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1},\hat{\boldsymbol{\Sigma}}_{\hat{z}\hat{z}}=\hat{\boldsymbol{\Sigma}}_{11}-\boldsymbol{B}\hat{\boldsymbol{\Sigma}}_{22}\boldsymbol{B}'$.

秩为 k 的 $\boldsymbol{\beta}$ 的极大似然估计量与正态分布的 $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{X}^{(2)}$ 相同, 这是因为 $\boldsymbol{X}=(\boldsymbol{X}^{(1)\prime},\boldsymbol{X}^{(2)\prime})'$ 的密度可分解成

$$n(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})=n(\boldsymbol{x}^{(1)}|\boldsymbol{\mu}^{(1)}+\boldsymbol{\beta}(\boldsymbol{x}^{(2)}-\bar{\boldsymbol{x}}^{(2)}),\boldsymbol{\Sigma}_{ZZ})n(\boldsymbol{x}^{(2)}|\boldsymbol{\mu}^{(2)},\boldsymbol{\Sigma}_{22}). \tag{3}$$

降秩回归已经应用到许多方面, 包括计量经济学、时间序列分析和信号处理. 比如, Johansen(1995) 将降秩回归应用到经济时间序列中协整的估计, Tsay and Tiao(1985) 和 Ahn and Reinsel (1988) 对平稳过程的应用, Stoica and Viberg(1996) 对信号处理的应用. 一般地, 估计的降秩回归在回归模型中是一个比无约束的估计量更好的估计量.

13.7 节在对最小二乘估计量 $\boldsymbol{B}=\hat{\boldsymbol{\Sigma}}_{12}\hat{\boldsymbol{\Sigma}}_{22}^{-1}$ 有足够的渐近正态性的假定下得到了降秩回归估计量的渐近分布. $\hat{\boldsymbol{B}}_k$ 的渐近分布由 Ryan, Hubert, Carter, Sprague, and Parrott(1992), Schmidli(1996), Stoica and Viberg(1996), Reinsel and Velu(1998) 在 $\boldsymbol{Z}_\alpha$ 服从正态分布的假定下用 Fisher 信息期望得到的. Izenman(1975) 提出了降秩回归这一术语.

12.8 联立方程模型

12.8.1 模型

计量经济学中的结构方程模型的推理同典型相关是有关的. 一般模型是

$$\boldsymbol{B}\boldsymbol{y}_t+\boldsymbol{\Gamma}\boldsymbol{z}_t=\boldsymbol{u}_t,\quad t=1,\cdots,T, \tag{1}$$

其中 $\boldsymbol{B}$ 是 $G\times G$, $\boldsymbol{\Gamma}$ 是 $G\times K$ 的. 这里 $\boldsymbol{y}_t$ 是由 G 个联合相依变量 (内生的) 组成的, $\boldsymbol{z}_t$ 是由被看作 "独立变量" 的 K 个预定变量 (外生的和滞后相依的) 组成的, $\boldsymbol{u}_i$ 由 G 个不可观测随机变量组成, 有

$$\mathrm{E}(\boldsymbol{u}_t)=\boldsymbol{0},\quad \mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t')=\boldsymbol{\Sigma}. \tag{2}$$

我们要求 $\boldsymbol{B}$ 是非奇异的. 这个模型由 Haavelmo(1944) 提出, 由 Koopmans, Marschak, Hurwicz, Anderson, Rubin, Leipnik, 等等, 于 1944-1954 在 Cowles 经济研究委员会推广. 每个分量方程表示某种群体的行为 (比如消费者和生产者), 有经济学的意义.

结构方程组 (1) 可以解出 $\boldsymbol{y}_t$, (因为 $\boldsymbol{B}$ 是非奇异的):

$$\boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{z}_t + \boldsymbol{\nu}_t, \tag{3}$$

其中

$$\boldsymbol{\Pi} = -\boldsymbol{B}^{-1}\boldsymbol{\Gamma}, \quad \boldsymbol{\nu}_t = \boldsymbol{B}^{-1}\boldsymbol{u}_t, \tag{4}$$

并且有

$$\mathrm{E}(\boldsymbol{\nu}_t) = \boldsymbol{0}, \quad \mathrm{E}\boldsymbol{\nu}_t\boldsymbol{\nu}_t' = \boldsymbol{B}^{-1}\boldsymbol{\Sigma}(\boldsymbol{B}')^{-1} = \boldsymbol{\Omega}, \tag{5}$$

将方程 (3) 称为模型的降秩形式. 它是一个多重回归模型. 原则上, 它是可观测的.

12.8.2 由给定零的识别

结构方程 (1) 可在左端乘一个任意的非奇异矩阵. 为了确定出具有经济学意义的分量方程, 必须加上约束条件. 例如, 在需求和供给的情况下, 描述需求的方程, 可以由它包括消费者收入而不包括原材料成本区别出, 后者在供给方程中. 后者的排除相当于令它在需求方程中的系数是 0.

我们考虑结构方程在给定某些系数为零后的识别. 考虑第一个方程比较简单. 假定变量是已编号的, 因此前 G_1 个联合相依变量包括在第一个方程中, 剩下的 $G_2 = G - G_1$ 个没有被包括, 前 K_1 个预定的变量被包括了, 剩下的 $K_2 = K - K_1$ 个被排除了. 我们可以将系数矩阵分成

$$\begin{pmatrix} \boldsymbol{B} & \boldsymbol{\Gamma} \end{pmatrix} = \begin{bmatrix} \boldsymbol{\beta}' & \boldsymbol{0} & \boldsymbol{\gamma}' & \boldsymbol{0} \\ - & - & - & - \end{bmatrix}, \tag{6}$$

其中向量 $\boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\gamma}, \boldsymbol{0}$ 分别有 G_1, G_2, K_1, K_2 个分量. 降秩形式被保形地分为 $\boldsymbol{G}_1$ 和 $\boldsymbol{G}_2$ 行组, $\boldsymbol{K}_1$ 和 $\boldsymbol{K}_2$ 列组:

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ - & - \end{bmatrix}. \tag{7}$$

$\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Pi}$ 的关系可以表示成

$$\begin{bmatrix} \boldsymbol{\gamma}' & \boldsymbol{0} \\ - & - \end{bmatrix} = \boldsymbol{\Gamma} = -\boldsymbol{B}\boldsymbol{\Pi} = -\begin{bmatrix} \boldsymbol{\beta}' & \boldsymbol{0} \\ - & - \end{bmatrix} \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ - & - \end{bmatrix} = -\begin{bmatrix} \boldsymbol{\beta}'\boldsymbol{\Pi}_{11} & \boldsymbol{\beta}\boldsymbol{\Pi}_{12} \\ - & - \end{bmatrix}. \tag{8}$$

(8) 式右上角的元产生

$$\boldsymbol{\beta}'\boldsymbol{\Pi}_{12} = \boldsymbol{0}. \tag{9}$$

为了除一个比例常数外唯一地确定 β $(G_1 \times 1)$, 我们需要

$$\operatorname{rank}(\boldsymbol{\Pi}_{12}) = G_1 - 1. \tag{10}$$

这就有

$$K_2 \geqslant G_1 - 1. \tag{11}$$

G_2 加上 (11) 式给出了*有序条件*

$$G_2 + K_2 \geqslant G_1 + G_2 - 1 = G - 1. \tag{12}$$

识别方程中给定的 0 的个数必须至少比方程 (或者联合相依变量) 的个数少 1.

可以证明, 当 $\boldsymbol{B}$ 非奇异时, (10) 式成立当且仅当矩阵的秩由 $(\boldsymbol{B}\ \boldsymbol{\Gamma})$ 的列构成, 其中在第一行中有 $G-1$ 个给定的零.

12.8.3 降秩形式的估计

模型 (3) 是典型的多重回归模型. 观测值是

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{z}_1 \end{pmatrix}, \cdots, \begin{pmatrix} \boldsymbol{y}_T \\ \boldsymbol{z}_T \end{pmatrix}. \tag{13}$$

$\boldsymbol{\Pi}$ 和 $\boldsymbol{\Omega}$ 通常的估计量 (8.2 节) 是

$$\boldsymbol{P} = \sum_{t=1}^{T} y_t z_t' \left(\sum_{t=1}^{T} z_t z_t' \right)^{-1}, \tag{14}$$

$$\hat{\boldsymbol{\Omega}} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \boldsymbol{P} z_t)(y_t - \boldsymbol{P} z_t)'. \tag{15}$$

如果 $\boldsymbol{\nu}_t$ 是正态分布的, 则这些是极大似然估计量.

如果 z_t 是外生的 (不考虑正态性), 那么

$$\mathrm{E}\,(\mathrm{vec}\,\boldsymbol{P}) = \mathrm{vec}\boldsymbol{\Pi}, \quad \mathrm{Cov}(\mathrm{vec}\,\boldsymbol{P}) = \boldsymbol{A}^{-1} \otimes \boldsymbol{\Omega}, \tag{16}$$

其中

$$\boldsymbol{A} = \sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z'}_t, \tag{17}$$

$\mathrm{vec}(\boldsymbol{d}_1, \cdots, \boldsymbol{d}_m) = (\boldsymbol{d'}_1, \cdots, \boldsymbol{d'}_m)'$. 如果更进一步有, $\boldsymbol{\nu}_t$ 是正态的, 那么 $\boldsymbol{P}$ 是正态分布的, $T\hat{\boldsymbol{\Omega}}$ 有 Wishart 分布, 协方差阵是 $\boldsymbol{\Omega}$, 自由度是 $T-K$.

12.8.4 方程系数的估计

首先, 考虑当 $K_2 = G_1 - 1$ 时 $\boldsymbol{\beta}$ 的系数向量的估计. 令

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{pmatrix} \tag{18}$$

如 $\boldsymbol{\Pi}$ 划分. 则以概率 1 有, $\mathrm{rank}(\boldsymbol{P}_{12}) = G_1 - 1$, 方程

$$\hat{\beta}' \boldsymbol{P}_{12} = \boldsymbol{0} \tag{19}$$

有非平凡的解, 并且解在除一个比例常数外唯一. 当误差项是正态分布时这即是极大似然估计量.

如果 $K_2 \geqslant G_1$, 那么以概率 1 有, $\mathrm{rank}(\boldsymbol{P}_{12}) = G_1$, (19) 式只有平凡的解 $\hat{\boldsymbol{\beta}} = \boldsymbol{0}$, 这是不满足要求的. 为了寻找合适的估计量, 我们令 $\hat{\beta}$ 在同另一个 $\hat{\beta}'$ 的函数相关的意义下极小化 $\boldsymbol{\beta}' P_{12}$.

令 $\boldsymbol{z}_t$ 分为 K_1 和 K_2 两部分子向量.

$$\boldsymbol{z}_t = \begin{pmatrix} \boldsymbol{z}_t^{(1)} \\ \boldsymbol{z}_t^{(2)} \end{pmatrix}, \tag{20}$$

$$\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t' = \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}, \tag{21}$$

$$\boldsymbol{A}_{22.1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}. \tag{22}$$

令 $\boldsymbol{y}_t$ 和 $\boldsymbol{\Omega}$ 分成 G_1 和 G_2 部分.

$$\boldsymbol{y}_t = \begin{pmatrix} \boldsymbol{y}_t^{(1)} \\ \boldsymbol{y}_t^{(2)} \end{pmatrix}, \tag{23}$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}. \tag{24}$$

现在我们建立起 $y^{(1)}$ 的多元方差分析表:

来源	平方和
$\boldsymbol{z}_t^{(1)}$	$\sum_{s,t=1}^{T} \boldsymbol{y}_s^{(1)} \boldsymbol{z}_s^{(1)\prime} \boldsymbol{A}_{11}^{-1} \boldsymbol{z}_t^{(1)} \boldsymbol{y}_t^{(1)\prime}$
$\boldsymbol{z}_t^{(2)} \perp \boldsymbol{z}_t^{(1)}$	$\boldsymbol{P}_{12}\boldsymbol{A}_{22.1}\boldsymbol{P}_{12}'$
误差	$\sum_{t=1}^{T}(\boldsymbol{y}_t^{(1)} - \boldsymbol{P}_{11}\boldsymbol{z}_t^{(1)} - \boldsymbol{P}_{12}\boldsymbol{z}_t^{(2)})(\boldsymbol{y}_t^{(1)} - \boldsymbol{P}_{11}\boldsymbol{z}_t^{(1)} - \boldsymbol{P}_{12}'\boldsymbol{z}_t^{(2)})'$
总和	$\sum_{t=1}^{T} \boldsymbol{y}_t^{(1)} \boldsymbol{y}_t^{(1)\prime}$

表中的第一项是由效应 $\boldsymbol{z}_t^{(1)}$ 导致的 $\boldsymbol{y}_t^{(1)}$ 的 (向量) 平方和. 第二项是由效应 $\boldsymbol{z}_t^{(2)}$ 在 $\boldsymbol{z}_t^{(1)}$ 之外的效应引起的. 两项加上 $(\boldsymbol{PAP}')_{11}$, 即 $\boldsymbol{z}_t$ 的总效应, 预定的变量.

我们建议找到向量 $\hat{\boldsymbol{\beta}}$, 使 $\boldsymbol{z}_t^{(2)}$ 的效应与 $\hat{\boldsymbol{\beta}}'\boldsymbol{y}_t^{(1)}$ 在 $\boldsymbol{z}_t^{(1)}$ 之外的效应相对于 $\hat{\boldsymbol{\beta}}'\boldsymbol{y}_t^{(1)}$ 的误差平方和极小化. 我们极小化

$$\frac{\hat{\boldsymbol{\beta}}'(\boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}_{12}')\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Omega}}_{11}\hat{\boldsymbol{\beta}}} = \frac{(\hat{\boldsymbol{\beta}}'\boldsymbol{P}_{12})\boldsymbol{S}_{22.1}(\hat{\boldsymbol{\beta}}'\boldsymbol{P}_{12})'}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Omega}}_{11}\hat{\boldsymbol{\beta}}}, \tag{25}$$

其中 $T\hat{\boldsymbol{\Omega}} = \sum_{t=1}^{T} \boldsymbol{y}_t\boldsymbol{y}_t - \boldsymbol{PAP}'$. 估计量被称为**最小方差比估计量**. 在正态下和基于这个单独方程系数的 0 约束下, 估计量是极大似然, 称为**极限信息极大似然估计量**(LIML) [Anderson and Rubin(1949)].

(25) 式的极小化即是找到

$$|\, \boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}_{12}' - \lambda\hat{\boldsymbol{\Omega}}_{11} \,| = 0 \tag{26}$$

的最小的根, 设为 ν. 相应的向量满足

$$\boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}_{12}'\hat{\boldsymbol{\beta}} = \nu\hat{\boldsymbol{\Omega}}_{11}\hat{\boldsymbol{\beta}}. \tag{27}$$

向量根据一些规则可以正规化. 一个频繁用到的规则是令某项系数 (非零) 等于 1, 不妨设为第一个, $\hat{\beta}_1 = 1$. 如果我们写成

$$\boldsymbol{\beta} = \begin{pmatrix} 1 \\ \boldsymbol{\beta}^* \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ \hat{\boldsymbol{\beta}}^* \end{pmatrix}, \tag{28}$$

$$\boldsymbol{\Pi}_{12} = \begin{pmatrix} \boldsymbol{\pi}_{12} \\ \boldsymbol{\Pi}_{12}^* \end{pmatrix}, \quad \boldsymbol{P}_{12} = \begin{pmatrix} \boldsymbol{p}_{12} \\ \boldsymbol{P}_{12}^* \end{pmatrix}, \tag{29}$$

$$\hat{\boldsymbol{\Omega}}_{11} = \begin{pmatrix} \hat{\omega}_{11} & \hat{\boldsymbol{\omega}}_{(1)}' \\ \hat{\boldsymbol{\omega}}_{(1)} & \hat{\boldsymbol{\Omega}}_{11}^* \end{pmatrix}, \tag{30}$$

那 (27) 式可以被线性方程

$$(\boldsymbol{P}_{12}^*\boldsymbol{S}_{22.1}\boldsymbol{P}_{12}^{*}{}' - \nu\hat{\boldsymbol{\Omega}}_{11}^*)\hat{\boldsymbol{\beta}}^* = -(\boldsymbol{P}_{12}^*\boldsymbol{S}_{22.1}\boldsymbol{p}_{12}' - \nu\hat{\boldsymbol{\omega}}_{(1)}) \tag{31}$$

取代. (27) 的第一个分量方程被删去了, 因为它和其他的方程线性相关 [因为 ν 是 (26) 式的一个根].

12.8.5 同线性函数关系的关系

我们现在证明单个线性函数关系的模型 ($q = 1$) 与结构方程模型在 $G_2 = 0(\boldsymbol{y}_t^{(1)} = \boldsymbol{y}_t), \boldsymbol{z}_t^{(1)} \equiv 1(K_1 = 1)$ 的特殊情况下是等同的. 将两个模型写为

$$\boldsymbol{X}_{\alpha j} = \boldsymbol{\mu} + \boldsymbol{\nu}_\alpha + \boldsymbol{U}_{\alpha j}, \qquad \alpha = 1, \cdots, n, \qquad j = 1, \cdots, k, \tag{32}$$

其中

$$\sum_{\alpha=1}^{n} \boldsymbol{\nu}_\alpha = \boldsymbol{0}, \tag{33}$$

$$\boldsymbol{y}_t = \boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2\boldsymbol{z}_t^{(2)} + \boldsymbol{\nu}_t, \quad t = 1, \cdots, T, \tag{34}$$

其中 $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_1, \boldsymbol{\Pi}_2)$. 两个模型的对应是 $p \leftrightarrow G = G_1$,

$$\boldsymbol{X}_{\alpha j} \leftrightarrow \boldsymbol{y}_t, \quad \boldsymbol{U}_{\alpha j} \leftrightarrow \boldsymbol{\nu}_t, \tag{35}$$

$$(\alpha, j) \leftrightarrow t, \quad nk \leftrightarrow T, \tag{36}$$

$$\boldsymbol{\Psi} \leftrightarrow \boldsymbol{\Omega}, \quad \boldsymbol{\mu} \leftrightarrow \boldsymbol{\Pi}_1. \tag{37}$$

我们用哑变量为线性函数关系写出模型. 定义

$$\boldsymbol{s}_{\alpha j} = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow 第\ \alpha\ 个位置 \quad \alpha = 1, \cdots, n-1, \tag{38}$$

$$\boldsymbol{s}_{nj} = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}. \tag{39}$$

则有

$$\boldsymbol{\mu}+\boldsymbol{\nu}_\alpha=(\boldsymbol{\mu},\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_{n-1})\begin{pmatrix}1\\ \boldsymbol{s}_{\alpha j}\end{pmatrix},\quad \alpha=1,\cdots,n, \tag{40}$$

其中 j 可能是被抑制的. 注意到

$$\boldsymbol{\nu}_n=-(\boldsymbol{\nu}_1+\cdots+\boldsymbol{\nu}_{n-1}). \tag{41}$$

相应的是

$$\mathbf{1}\leftrightarrow \boldsymbol{z}_t^{(1)},\quad \boldsymbol{s}_{\alpha j}\leftrightarrow \boldsymbol{z}_t^{(2)}, \tag{42}$$

$$\boldsymbol{\mu}\leftrightarrow \boldsymbol{\Pi}_1,\qquad (\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_{n-1})\leftrightarrow \boldsymbol{\Pi}_2, \tag{43}$$

$$1\leftrightarrow \boldsymbol{K}_1,\quad n-1\leftrightarrow \boldsymbol{K}_2, \tag{44}$$

$$\boldsymbol{B}(\boldsymbol{\nu}_1,\cdots,\boldsymbol{\nu}_{n-1})=\mathbf{0}\leftrightarrow \boldsymbol{\beta}'\boldsymbol{\Pi}_2=0. \tag{45}$$

令 $\boldsymbol{P}=(\boldsymbol{P}_1\ \ \boldsymbol{P}_2)$. 用统计的方法我们有相应的

$$\hat{\boldsymbol{\mu}}=\bar{\boldsymbol{x}}\leftrightarrow \bar{\boldsymbol{y}}, \tag{46}$$

$$\hat{\boldsymbol{\nu}}_\alpha=\bar{\boldsymbol{x}}_\alpha-\bar{\boldsymbol{x}}\leftrightarrow \boldsymbol{P}_2. \tag{47}$$

效应矩阵是

$$\boldsymbol{H}=k\sum_{\alpha=1}^{n}(\bar{\boldsymbol{x}}_\alpha-\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_\alpha-\bar{\boldsymbol{x}})'\leftrightarrow \boldsymbol{P}_2\boldsymbol{A}_{22.1}\boldsymbol{P}_2', \tag{48}$$

误差矩阵是

$$\boldsymbol{G}=\sum_{\alpha=1}^{n}\sum_{j=1}^{k}(\boldsymbol{x}_{\alpha j}-\bar{\boldsymbol{x}}_\alpha)(\boldsymbol{x}_{\alpha j}-\bar{\boldsymbol{x}}_\alpha)'\leftrightarrow T\hat{\boldsymbol{\Omega}}=\sum_{t=1}^{T}(y_t-\boldsymbol{P}z_t)(y_t-\boldsymbol{P}z_t)'. \tag{49}$$

则 $q=1$ 时线性函数关系的估计量 $\hat{\boldsymbol{B}}$ 同 LIML 估计量是等同的. [Anderson (1951b), (1976), (1984a)].

12.8.6 $T\to\infty$ 时的渐近理论

我们可以得到由 (28) 式和 (31) 式定义的 $\sqrt{T}(\hat{\beta}^*-\beta^*)$ 的极限分布, 通过证明 $\hat{\beta}^*$ 渐近等价于

$$\hat{\boldsymbol{\beta}}^*_{\text{TSLS}}=-(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12})^{-1}\boldsymbol{P}^*_{22}\boldsymbol{S}_{22.1}\boldsymbol{p}'_{12}. \tag{50}$$

这个推导除了记号外本质上同 Anderson and Rubin (1950) 给出的相同. (50) 式定义的估计量, 称为*二级最小二乘*(TSLS) 估计量, 是对从 (31) 式剔除项 $\nu\hat{\boldsymbol{\Omega}}^*_{11}$ 和 $\nu\hat{\boldsymbol{\omega}}_{(1)}$ 得到的 LIML 估计量的逼近. 令 $\hat{\beta}^*=\hat{\beta}^*_{\text{LIML}}$. 我们假设 $\sqrt{T}(\boldsymbol{P}-\boldsymbol{\Pi})$ 的条件有极限正态分布. (见定理 8.11.1.)

引理 12.8.1 *假设当 $T\to\infty$ 时 $(1/T)\boldsymbol{A}\to\boldsymbol{A}^0$, A^0 是正定矩阵. 则 $\nu=O_p(1/T)$, 其中 ν 是 (26) 式的最小根.*

证明 令 $\tilde{\boldsymbol{P}}_{12}=\sqrt{T}(\boldsymbol{P}_{12}-\boldsymbol{\Pi}_{12})$. 则因为 $\boldsymbol{\beta}'\boldsymbol{\Pi}_{12}=\mathbf{0}$, 所以

$$\frac{\boldsymbol{\beta}'\boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}_{12}\boldsymbol{\beta}}{\boldsymbol{\beta}'\hat{\boldsymbol{\Omega}}_{11}\boldsymbol{\beta}}=\frac{\boldsymbol{\beta}'[\boldsymbol{\Pi}_{12}+(1/\sqrt{T})\tilde{\boldsymbol{P}}_{12}]\boldsymbol{S}_{22.1}[\boldsymbol{\Pi}_{12}+(1/\sqrt{T})\tilde{\boldsymbol{P}}_{12}]'\boldsymbol{\beta}}{\boldsymbol{\beta}'\hat{\boldsymbol{\Omega}}_{11}\boldsymbol{\beta}}$$

$$=\frac{\boldsymbol{\beta}'\tilde{\boldsymbol{P}}_{12}\boldsymbol{S}_{22.1}\tilde{\boldsymbol{P}}'_{12}\boldsymbol{\beta}}{T\boldsymbol{\beta}'\hat{\boldsymbol{\Omega}}_{11}\boldsymbol{\beta}}=\boldsymbol{O}_p\left(\frac{1}{T}\right). \tag{51}$$

由于

$$\nu=\min_{\hat{\boldsymbol{\beta}}}\frac{\hat{\boldsymbol{\beta}}'\boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}'_{12}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Omega}}_{11}\hat{\boldsymbol{\beta}}}\leqslant\frac{\boldsymbol{\beta}'\boldsymbol{P}_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}'_{12}\boldsymbol{\beta}}{\boldsymbol{\beta}'\hat{\boldsymbol{\Omega}}_{11}\boldsymbol{\beta}}, \tag{52}$$

引理得证. ■

引理 12.8.1 的意义在于指出 LIML 估计量和 TSLS 估计量之差是 $O_p(1/T)$. 我们有

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}^*_{\text{LIML}}-\hat{\boldsymbol{\beta}}^*_{\text{TSLS}} \\
=&(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12})^{-1}\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{p}'_{12} \\
&-(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12}-\nu\hat{\boldsymbol{\Omega}}_{11})^{-1}(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{p}'_{12}-\nu\hat{\boldsymbol{\omega}}_{(1)}) \\
=&[(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12})^{-1}-(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12}-\nu\hat{\boldsymbol{\Omega}}_{11})^{-1}]\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{p}'_{12} \\
&+(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12}-\nu\hat{\boldsymbol{\Omega}}_{11})^{-1}\nu\hat{\boldsymbol{\omega}}_{(1)} \\
=&-\nu(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12})^{-1}\hat{\boldsymbol{\Omega}}_{11}(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^{*\prime}_{12}-\nu\hat{\boldsymbol{\Omega}}_{11})^{-1}\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{p}'_{12} \\
&+\nu(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^*_{12}-\nu\hat{\boldsymbol{\Omega}}_{11})^{-1}\hat{\boldsymbol{\omega}}_{(1)} \\
=&O_p(\nu)=O_p\left(\frac{1}{T}\right).
\end{aligned} \tag{53}$$

考虑

$$\boldsymbol{p}'_{12}+\boldsymbol{P}^{*\prime}_{12}\boldsymbol{\beta}^*=\boldsymbol{P}'_{12}\boldsymbol{\beta}=\boldsymbol{A}^{-1}_{22.1}\sum_{t=1}^{T}\boldsymbol{z}^{(2\cdot1)}_t\boldsymbol{y}^{(1)}_t\boldsymbol{\beta}=\boldsymbol{A}^{-1}_{22.1}\sum_{t=1}^{T}\boldsymbol{z}^{(2\cdot1)}_t u_{1t}, \tag{54}$$

其中 $\boldsymbol{z}^{(2\cdot1)}_t=\boldsymbol{z}^{(2)}_t-\boldsymbol{A}_{21}\boldsymbol{A}^{-1}_{11}\boldsymbol{z}^{(1)}_t$. 因此 $\mathrm{E}(\boldsymbol{p}'_{12}+\boldsymbol{P}^{*\prime}_{12}\boldsymbol{\beta}')=\mathrm{E}(\boldsymbol{P}'_{12}\boldsymbol{\beta})=\boldsymbol{0}$, 还有

$$\mathrm{E}(\boldsymbol{p}'_{12}+\boldsymbol{P}^{*\prime}_{12}\boldsymbol{\beta}^*)(\boldsymbol{p}'_{12}+\boldsymbol{P}^{*\prime}_{12}\boldsymbol{\beta}^*)'=\mathrm{E}(\boldsymbol{P}'_{12}\boldsymbol{\beta}(\boldsymbol{P}'_{12}\boldsymbol{\beta})')=\sigma_{11}\boldsymbol{A}^{-1}_{22.1}. \tag{55}$$

注意到 $\hat{\boldsymbol{\beta}}^*_{\text{TSLS}}-\boldsymbol{\beta}^*=-(\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}^*_{12})^{-1}\boldsymbol{P}^*_{12}\boldsymbol{S}_{22.1}\boldsymbol{P}'_{12}\boldsymbol{\beta}$ 和 $(\boldsymbol{\beta}',\boldsymbol{0})\boldsymbol{y}_t+(\boldsymbol{\gamma}',\boldsymbol{0})\boldsymbol{z}_t=\boldsymbol{u}_{1t}$.

定理 12.8.1 *在定理 8.11.1 的条件下*

$$\sqrt{T}(\hat{\boldsymbol{\beta}}^*_{\text{LIML}}-\boldsymbol{\beta}^*)\xrightarrow{d}N[\boldsymbol{0},\sigma_{11}(\boldsymbol{\Pi}_{12}\boldsymbol{S}^0_{22.1}\boldsymbol{\Pi}'_{12})^{-1}]. \tag{56}$$

证明 由 (55) 式就可证明定理, $\boldsymbol{S}_{22.1}\to\boldsymbol{S}^0_{22.1}$, $\boldsymbol{P}_{12}\xrightarrow{p}\boldsymbol{\Pi}_{12}$. ■

由于 LIML 估计量与 12.7.5 节所提到的线性函数关系的极大似然估计量之间的对应, 这个渐近理论也可以应用到后者. 假设单个线性函数关系可写成

$$0=\boldsymbol{\beta}'\boldsymbol{\nu}_\alpha=(1\quad\boldsymbol{\beta}^*)\begin{pmatrix}\nu_{1\alpha}\\ \boldsymbol{\nu}^*_\alpha\end{pmatrix}=\nu_{1\alpha}+\boldsymbol{\beta}^{*\prime}\boldsymbol{\nu}^*_\alpha,\quad\alpha=1,\cdots,n, \tag{57}$$

其中

$$\boldsymbol{\nu}_\alpha = \begin{pmatrix} \nu_{1\alpha} \\ \boldsymbol{\nu}_\alpha^* \end{pmatrix}, \quad \alpha = 1, \cdots, n. \tag{58}$$

令 $n(\leftrightarrow K)$ 是固定的, 令重复数 $k \to \infty$(对固定的 K 相应于 $T/K \to \infty$). 令 $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Psi}\boldsymbol{\beta}$.

由于 $\boldsymbol{\Pi}_{12}\boldsymbol{A}_{22.1}\boldsymbol{\Pi}_{12}'$ 对应于 $k\sum_{\alpha=1}^n \boldsymbol{\nu}_\alpha^*\boldsymbol{\nu}_\alpha^{*\prime}$, 这里 $\hat{\boldsymbol{\beta}}^*$ 有渐近分布

$$N\left[\boldsymbol{\beta}^*, \sigma^2\left(k\sum_{\alpha=1}^n \boldsymbol{\nu}_\alpha^*\boldsymbol{\nu}_\alpha^{*\prime}\right)^{-1}\right]. \tag{59}$$

虽然 Anderson and Rubin(1950) 证明了 $\nu\hat{\boldsymbol{\Omega}}_{11}^*$ 和 $\nu\hat{\boldsymbol{\omega}}_{(1)}$ 可以从 (31) 式定义的 $\hat{\boldsymbol{\beta}}_{\text{LIML}}^*$ 中剔除, 因此 $\hat{\boldsymbol{\beta}}_{\text{LIML}}^*$ 渐近等价于 $\hat{\boldsymbol{\beta}}_{\text{TSLS}}^*$, 但他们没有显式给出 $\hat{\boldsymbol{\beta}}_{\text{TSLS}}^*$. [作为 Cowles 委员会项目的一部分, Chernoff and Divinsky(1953) 得到了 $\hat{\boldsymbol{\beta}}_{\text{LIML}}$ 的计算程序.] 该 TSLS 估计量被 Basmann(1957) 和 Theil(1961) 提出. 它相应于在第一坐标下对线性函数关系建立普通最小二乘. 如果 $\boldsymbol{\beta}$ 的其他系数等于零, 最小值将在那个坐标的方向.

当误差协方差阵是未知的, 并且有重复时, 考虑一般线性函数关系. 限制 $\boldsymbol{B}$ 为

$$\boldsymbol{B} = (\boldsymbol{I}_m \quad \boldsymbol{B}^*). \tag{60}$$

分成

$$\boldsymbol{\nu}_\alpha = \begin{pmatrix} \boldsymbol{\nu}_\alpha^{(1)} \\ \boldsymbol{\nu}_\alpha^{(2)} \end{pmatrix}. \tag{61}$$

则 $\boldsymbol{B}^*$ 的最小二乘估计量是

$$\hat{\boldsymbol{B}}_{\text{LS}}^* = -\sum_{\alpha=1}^n (\bar{\boldsymbol{x}}_\alpha^{(1)} - \bar{\boldsymbol{x}}^{(1)})(\bar{\boldsymbol{x}}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})' \left[\sum_{\alpha=1}^n (\bar{\boldsymbol{x}}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})(\bar{\boldsymbol{x}}_\alpha^{(2)} - \bar{\boldsymbol{x}}^{(2)})'\right]^{-1}. \tag{62}$$

对固定的 n 和 $k \to \infty$, $\hat{\boldsymbol{B}}_{\text{LS}}^* \xrightarrow{p} \boldsymbol{B}^*$,

$$\sqrt{k}\text{vec}(\hat{\boldsymbol{B}}_{\text{LS}}^* - \boldsymbol{B}^*) \to N\left[\boldsymbol{0}, \left(\sum_{\alpha=1}^n \boldsymbol{\nu}_\alpha^{(2)}\boldsymbol{\nu}_\alpha^{(2)\prime}\right)^{-1} \otimes \boldsymbol{B}\boldsymbol{\Psi}\boldsymbol{B}'\right]. \tag{63}$$

[见 Anderson(1984b).]Anderson(1951c) 证明了 q 个最小的样本根有极大似然估计量是渐近等价的概率顺序, 即 $\sqrt{k}\text{vec}(\hat{\boldsymbol{B}}_{\text{LS}}^* - \boldsymbol{B}^*)$ 的极限分布是 (63) 式的右端.

12.8.7 其他渐近理论

用线性函数关系的方法, 考虑 $n \to \infty$ 和 k 是固定的可能更自然. 当 $k = 1$ 和误差协方差阵是 $\sigma^2\boldsymbol{I}_p$ 时, Gleser(1981) 给出了渐近理论. 对联立方程模型, 相应的条件是 $K_2 \to \infty, T \to \infty$, K_2/T 逼近一个正极限. Kunitomo(1980) 给出了一个当 $p = 2, m = q = 1$ 时分布的渐近展开.

当 $n \to \infty$ 时, 最小二乘估计量即在某个确定的方向上极小化残差平方和) 是不相容的; LIML 和 TSLS 估计量不是渐近等价的.

12.8.8　估计量的分布

计量经济学家集中研究了 TSLS 和 LIML 估计量的分布, 尤其是两组内生变量的情况.

Basmann(1961), (1963), Richardson(1968), Sawa(1969), Mariano and Sawa(1972), Phillips(1980), 和 Anderson and Sawa(1982) 给出了精确分布. 这些并不是特别有价值, 因为它们经常用具有未知或者不相干的性质的无穷序列的方法给出.

一个有用的方法是近似分布. 分布的渐近展开由 Sargan and Mikhail(1971), Anderson and Sawa(1973), Anderson(1974), Kunitomo(1980) 以及其他文献给出. Phillips(1982) 研究了 Padé 近似. 见 Anderson(1977).

TSLS 和 LIML 估计量在两个内生变量的情况下的分布表由 Anderson and Sawa(1977), (1979), 和 Anderson, Kunitomo, and Sawa(19831) 给出.

Anderson, Kunitomo, and Sawa(1983b) 绘出线性函数关系 (12.6 节) 在 $p=2, m=q=1, \boldsymbol{\Psi}=\sigma^2\boldsymbol{\Psi}_0$ 以及不同的 β, n 和

$$\delta^2=\frac{1}{\sigma^2}\sum_{\sigma=1}^{n}(\mu_\sigma-\bar{\mu})^2 \tag{64}$$

的情况下的极大似然估计量和最小二乘估计量 (某方向极小化) 的密度.

习　题

12.1　(12.2 节)　令 $z_\alpha=z_{1\alpha}=1, \alpha=1,\cdots,n, \boldsymbol{\beta}=\beta$. 证明 $\boldsymbol{\alpha}^{(1)}=\boldsymbol{\Sigma}^{-1}\beta$. 将这个结果同判别式函数建立联系 (第 6 章).

12.2　证明 (14) 式的根是实数.

12.3　(12.2 节)

(a) 令 $\boldsymbol{X}'=(\boldsymbol{X}^{(1)'}\ \boldsymbol{X}^{(2)'}), \mathrm{E}(\boldsymbol{X})=0$,

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{X}')=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix},$$

$U=\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}, V=\boldsymbol{\gamma}'\boldsymbol{X}^{(2)}, \mathrm{E}(U^2)=1=\mathrm{E}(V^2)$, 其中 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 是向量. 证明选择 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 极大化 $\mathrm{E}(UV)$ 等价于选择 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 极小化 $(U\ V)$ 的广义方差.

(b) 令 $\boldsymbol{X}'=(\boldsymbol{X}^{(1)'}\ \boldsymbol{X}^{(2)'}\ \boldsymbol{X}^{(3)'}), \mathrm{E}(\boldsymbol{X})=0$,

$$\mathrm{E}(\boldsymbol{X}\boldsymbol{X}')=\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23}\\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33}\end{pmatrix},$$

$U=\boldsymbol{\alpha}'\boldsymbol{X}^{(1)}, V=\boldsymbol{\gamma}'\boldsymbol{X}^{(2)}, W=\boldsymbol{\beta}'\boldsymbol{X}^{(3)}, \mathrm{E}(U^2)=\mathrm{E}(V^2)=\mathrm{E}(W^2)=1$. 考虑寻找 $\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\beta}$ 极小化 (U,V,W) 的广义方差. 证明这个极小值是关于变换 $\boldsymbol{X}^{*(i)}=\boldsymbol{A}_i\boldsymbol{X}^i(|\boldsymbol{A}_i|\neq 0)$ 的不变量.

(c) 用这个变换, 变换 $\boldsymbol{\Sigma}$ 到最简单的可能形式.

(d) 在 $\boldsymbol{X}^{(i)}$ 由两个元组成的情况下, 将问题简化 (极小化广义方差) 到最简单的形式.

(e) 给出这种情况下的导数方程.

(f) 证明极小化广义方差是 1 当且仅当 $\boldsymbol{\Sigma}_{12}=0, \boldsymbol{\Sigma}_{13}=0, \boldsymbol{\Sigma}_{23}=0$. (注: 这个典型变量概念的拓展并不适合于 "完美的" 显式处理.)

12.4 (见 12.2) 令

$$\boldsymbol{X}^{(1)}=\boldsymbol{A}\boldsymbol{Z}+\boldsymbol{Y}^{(1)},$$

$$\boldsymbol{X}^{(2)}=\boldsymbol{B}\boldsymbol{Z}+\boldsymbol{Y}^{(2)},$$

其中 $\boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(2)}, \boldsymbol{Z}$ 是独立的, 均值为零, 协方差阵 $\boldsymbol{I}$ 有适当的维数. 令 $\boldsymbol{A}=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_k), \boldsymbol{B}=(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_k)$, 假设 $\boldsymbol{A}'\boldsymbol{A}, \boldsymbol{B}'\boldsymbol{B}$ 是有正对角元的对角矩阵. 证明有非零典型相关的典型变量同 $\boldsymbol{a}_i'\boldsymbol{X}^{(1)}, \boldsymbol{b}_i'\boldsymbol{X}^{(2)}$ 是成比例的. 并求典型相关系数和适当的典型变量的正规化系数.

12.5 (12.2 节) 令 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_q > 0$ 是 (14) 式的正根, 其中 $\boldsymbol{\Sigma}_{11}$ 和 $\boldsymbol{\Sigma}_{22}$ 是 $q\times q$ 非奇异矩阵.

(a) $\boldsymbol{\Sigma}_{12}$ 的秩是什么?

(b) 将 $\Pi_{i=1}^q \lambda_i^2$ 写成 $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{21}$ 和 $\boldsymbol{\Sigma}_{22}$ 的有理函数的行列式. 验证你的结论.

(c) 如果 $\lambda_q=1$,

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

的秩是多少?

12.6 (12.2 节) 令 $\boldsymbol{\Sigma}_{11}=(1-g)\boldsymbol{I}_{p_1}+g\boldsymbol{\varepsilon}_{p_1}\boldsymbol{\varepsilon}_{p_1}', \boldsymbol{\Sigma}_{22}=(1-h)\boldsymbol{I}_{p_2}+h\boldsymbol{\varepsilon}_{p_2}\boldsymbol{\varepsilon}_{p_2}', \boldsymbol{\Sigma}_{12}=k\boldsymbol{\varepsilon}_{p_1}\boldsymbol{\varepsilon}_{p_2}'$, 其中 $-1/(p_1-1)<g<1, -1/(p_2-1)<h<1$, k 有恰当的限制. 求典型相关和典型变量. 对 k 的恰当的约束是什么?

12.7 (12.3 节) 求习题 4.42 中前两个变量和最后三个变量的典型相关和典型变量.

12.8 (12.3 节) 直接证明定理 12.2.1 的样本模拟.

12.9 (12.3 节) 如果 $\boldsymbol{\alpha}(0)$ 满足 $\boldsymbol{\alpha}'(0)\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}^{(1)}\neq 0$, 证明 $\lambda_1^2(i+1)\to\lambda_1^2$ 和 $\boldsymbol{\alpha}(i+1)\to\boldsymbol{\alpha}^{(1)}$. [提示: 利用 $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}=\boldsymbol{A}\boldsymbol{\Lambda}^2\boldsymbol{A}^{-1}$.]

12.10 (12.6 节) 证明 (9), (10), (11).

12.11 令 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q$ 是 $|\boldsymbol{\Sigma}_1-\lambda\boldsymbol{\Sigma}_2|=0$ 的根, 其中 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$ 是 $q\times q$ 维正定协方差阵.

(a) $\lambda_1=\lambda_q=1$ 包含了 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$ 之间的什么关系?

(b) $\lambda_q>1$ 包含了椭球面 $\boldsymbol{x}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x}=\boldsymbol{c}$ 和 $\boldsymbol{x}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x}=\boldsymbol{c}$ 的什么关系?

(c) $\lambda_1>1$ 和 $\lambda_q<1$ 包含了椭球面 $\boldsymbol{x}'\boldsymbol{\Sigma}_1^{-1}\boldsymbol{x}=\boldsymbol{c}$ 和 $\boldsymbol{x}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{x}=\boldsymbol{c}$ 之间的什么关系?

12.12 (12.4 节) 对 $q=2$ 用典型相关的方法给出 9.5 节的准则 (2).

12.13 对习题 9.11 的数据给出典型相关.

第 13 章　特征根和特征向量的分布

13.1　引　　言

本章我们得到当所有的总体方差是 1 时, 样本主成分向量的分布和它们的样本方差的分布 (13.3 节). 我们同样得到样本典型相关的分布, 当两组初始变量是独立时得到一组典型向量. 将会证明第二个分布等价于下一节介绍的根和向量的分布. 因为许多不变检验是特征根的函数, 所以我们对根的分布有特殊的兴趣. 举例说, 一般线性假设的不变检验 (8.6 节) 仅仅通过行列式方程

$$|(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)\boldsymbol{A}_{11.2}(\hat{\boldsymbol{\beta}}_{1\Omega}-\boldsymbol{\beta}_1^*)'-lN\hat{\boldsymbol{\Sigma}}_\Omega|=0 \tag{1}$$

的根依赖于样本. 如果假设为真, 那么特征根有由定理 13.2.2 或者定理 13.2.3 给出的分布. 因此一般线性假设的任何不变检验的显著性水平可由下节导出的分布获得. 如果检验准则是一个有序的根 (例如, 最大的根), 那么期望分布是特征根的联合分布的边缘分布.

特征根的极限分布在完全一般的条件下获得. 需要这些来得到其他的极限分布, 比如检验主成分的最小方差是相等的准则的分布. 一些极限分布由椭球等高分布得到.

13.2　两个 Wishart 矩阵的情况

13.2.1　变换

我们考虑 $\boldsymbol{A}^*$ 和 $\boldsymbol{B}^*$ $(p\times p)$ 独立地服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)(m,n\geqslant p)$ 分布. 我们称

$$|\boldsymbol{A}^*-l\boldsymbol{B}^*|=0 \tag{1}$$

的根为$\boldsymbol{A}^*$ 在 $\boldsymbol{B}^*$ 的度量下的特征根, 称满足

$$(\boldsymbol{A}^*-l\boldsymbol{B}^*)\boldsymbol{x}^*=\boldsymbol{0} \tag{2}$$

的向量为$\boldsymbol{A}^*$ 在 $\boldsymbol{B}^*$ 度量下的特征向量. 本节我们考虑这些根和向量的分布. 后面将会证明, 如果总体典型相关均为零, 则典型相关系数的平方有这个分布.

首先变换 $\boldsymbol{A}^*$ 和 $\boldsymbol{B}^*$ 使得分布不牵涉一个任意的矩阵 $\boldsymbol{\Sigma}$. 令 $\boldsymbol{C}$ 是满足 $\boldsymbol{C\Sigma C}'=\boldsymbol{I}$ 的矩阵. 令

$$\boldsymbol{A}=\boldsymbol{CA}^*\boldsymbol{C}',\quad \boldsymbol{B}=\boldsymbol{CB}^*\boldsymbol{C}'. \tag{3}$$

则 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别独立地服从 $W(\boldsymbol{I},m)$ 和 $W(\boldsymbol{I},n)$ 分布 (7.3.3 节). 因为

$$|\boldsymbol{A}-l\boldsymbol{B}|=|\boldsymbol{C}\boldsymbol{A}^*\boldsymbol{C}'-l\boldsymbol{C}\boldsymbol{B}^*\boldsymbol{C}'|=|\boldsymbol{C}(\boldsymbol{A}^*-l\boldsymbol{B}^*)\boldsymbol{C}'|=|\boldsymbol{C}|\cdot|\boldsymbol{A}^*-l\boldsymbol{B}^*|\cdot|\boldsymbol{C}'|,$$

所以 (1) 式的根正是

$$|\boldsymbol{A}-l\boldsymbol{B}|=0 \tag{4}$$

的根. 满足

$$(\boldsymbol{A}-l\boldsymbol{B})\boldsymbol{x}=\boldsymbol{0} \tag{5}$$

的相应的向量满足

$$\begin{aligned}\boldsymbol{0}&=\boldsymbol{C}^{-1}(\boldsymbol{A}-l\boldsymbol{B})\boldsymbol{x}\\&=\boldsymbol{C}^{-1}(\boldsymbol{C}\boldsymbol{A}^*\boldsymbol{C}'-l\boldsymbol{C}\boldsymbol{B}^*\boldsymbol{C}')\boldsymbol{x}\\&=(\boldsymbol{A}^*-l\boldsymbol{B}^*)\boldsymbol{C}'\boldsymbol{x}.\end{aligned} \tag{6}$$

因此向量 $\boldsymbol{x}^*$ 正是向量 $\boldsymbol{C}'x$.

考虑

$$|\boldsymbol{A}-f(\boldsymbol{A}+\boldsymbol{B})|=0 \tag{7}$$

的根以及满足

$$[\boldsymbol{A}-f(\boldsymbol{A}+\boldsymbol{B})]\,\boldsymbol{y}=\boldsymbol{0}. \tag{8}$$

的向量 y 将会更方便, 后一方程可写成

$$\boldsymbol{0}=(\boldsymbol{A}-f\boldsymbol{A}-f\boldsymbol{B})\boldsymbol{y}=[(1-f)\boldsymbol{A}-f\boldsymbol{B}]\boldsymbol{y}. \tag{9}$$

因为 $f=1$(例如 $|-\boldsymbol{B}|=0$) 的概率是 0, 上面的方程是

$$\left(\boldsymbol{A}-\frac{f}{1-f}\boldsymbol{B}\right)\boldsymbol{y}=\boldsymbol{0}. \tag{10}$$

因此 (4) 式的根是与 (7) 式的根有关的, 有 $l=f/(1-f)$ 或者 $f=l/(1+l)$, 满足 (5) 式的向量同满足 (8) 式的那些向量是等同的 (或者成比例的).

我们考虑求满足 (7) 和 (8) 的根和向量的分布. 令根有顺序 $f_1>f_2>\cdots>f_p>0$, 由于两个根相等的概率是 0[Okamoto(1973)]. 令

$$\boldsymbol{F}=\begin{pmatrix}f_1&0&\cdots&0\\0&f_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&f_p\end{pmatrix}. \tag{11}$$

假设 (8) 式由

$$\boldsymbol{y}'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{y}=1 \tag{12}$$

正规化后的相应的向量解是 $y_1,\cdots,y_p$. 这些向量必须满足

$$\boldsymbol{y}_i'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{y}_j=0, \tag{13}$$

因为只有 $\boldsymbol{y}_i'\boldsymbol{A}\boldsymbol{y}_j=f_j\boldsymbol{y}_i'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{y}_j$, $\boldsymbol{y}_i'\boldsymbol{A}\boldsymbol{y}_j=f_i\boldsymbol{y}_i'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{y}_j$ 才能使 (13) 式成立 $(f_i\neq f_j)$.

令 $p\times p$ 矩阵 $\boldsymbol{Y}$ 是

$$\boldsymbol{Y}=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_p). \tag{14}$$

方程 (8) 可以概括成

$$\boldsymbol{AY} = (\boldsymbol{A}+\boldsymbol{B})\boldsymbol{YF}, \tag{15}$$

(12) 和 (13) 给出

$$\boldsymbol{Y}'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{Y} = \boldsymbol{I}. \tag{16}$$

从 (15) 式我们得到

$$\boldsymbol{Y}'\boldsymbol{AY} = \boldsymbol{Y}'(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{YF} = \boldsymbol{F}. \tag{17}$$

将 (16) 式和 (17) 式左乘 $(\boldsymbol{Y}')^{-1}$ 右乘 $\boldsymbol{Y}^{-1}$, 得到

$$\begin{aligned}\boldsymbol{A}+\boldsymbol{B} &= (\boldsymbol{Y}')^{-1}\boldsymbol{Y}^{-1},\\ \boldsymbol{A} &= (\boldsymbol{Y}')^{-1}\boldsymbol{FY}^{-1}.\end{aligned} \tag{18}$$

现在令 $\boldsymbol{Y}^{-1} = \boldsymbol{E}$. 那么

$$\begin{aligned}\boldsymbol{A}+\boldsymbol{B} &= \boldsymbol{E}'\boldsymbol{E},\\ \boldsymbol{A} &= \boldsymbol{E}'\boldsymbol{FE},\\ \boldsymbol{B} &= \boldsymbol{E}'(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{E}.\end{aligned} \tag{19}$$

现在我们考虑 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 的联合分布. 从 (19) 式我们看出 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 唯一定义了 $\boldsymbol{A}$ 和 $\boldsymbol{B}$. 从 (7) 式和 (11) 式以及顺序量 $f_1 > \cdots > f_p$ 我们看出 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 唯一定义了 $\boldsymbol{F}$. $f = f_i$ 时的方程 (8) 和 (12) 除了 -1 的倍数外唯一定义了 $\boldsymbol{y}_i$(例如, 用 $-\boldsymbol{y}_i$ 代替 $\boldsymbol{y}_i$). 因为 $\boldsymbol{YE} = \boldsymbol{I}$, 这意味着 $\boldsymbol{E}$ 是唯一定义的, 除了 $\boldsymbol{E}$ 的行被 -1 乘外. 为了消除这种不确定性, 我们需要 $e_{i1} \geqslant 0$. ($e_{i1} = 0$ 的概率是 0.) 因此用 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 定义的 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是唯一的.

13.2.2 雅可比行列式

为了求 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 的密度, 我们根据 (19) 式用 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的密度代替, 再乘变换的雅可比行列式. 本小节我们致力于求得雅可比行列式

$$\left|\frac{\partial(\boldsymbol{A},\boldsymbol{B})}{\partial(\boldsymbol{E},\boldsymbol{F})}\right|. \tag{20}$$

因为从 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 到 $\boldsymbol{A}$ 和 $\boldsymbol{G} = \boldsymbol{A}+\boldsymbol{B}$ 的变换有单位雅可比行列式, 所以我们将求得

$$\left|\frac{\partial(\boldsymbol{A},\boldsymbol{G})}{\partial(\boldsymbol{E},\boldsymbol{F})}\right| = \left|\frac{\partial(\boldsymbol{A},\boldsymbol{B})}{\partial(\boldsymbol{E},\boldsymbol{F})}\right|. \tag{21}$$

首先我们注意到, 如果 $x_\alpha = f_\alpha(y_1,\cdots,y_n), \alpha = 1,\cdots,n$, 是一个一到一的变换, 那么雅可比行列式是线性变换

$$\mathrm{d}x_\alpha = \sum_\beta \frac{\partial f_\alpha}{\partial y_\beta}\mathrm{d}y_\beta \tag{22}$$

的行列式, 其中 $\mathrm{d}x_\alpha$ 和 $\mathrm{d}y_\beta$ 仅仅是形式微分 (即我们这样写有助于记忆). 如果 $f_\alpha(y_1,\cdots,y_n)$ 是多项式, 那么 $\partial f_\alpha/\partial y_\beta$ 是 $f_\alpha(y_1+y_1^*,\cdots,y_n+y_n^*)$ 的展开式中 y_β^* 的系数 [实际上是 $f_\alpha(y_1,\cdots,y_{\beta-1},y_\beta+y_\beta^*,y_{\beta+1},\cdots,y_n)$ 的展开式中的系数]. $\boldsymbol{A}$ 和 $\boldsymbol{G}$ 的元是 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 的多项式. 因此 $\boldsymbol{A}$ 中元的导数是 $(\boldsymbol{E}+\boldsymbol{E}^*)'(\boldsymbol{F}+\boldsymbol{F}^*)(\boldsymbol{E}+\boldsymbol{E}^*)$

的展开式中 $\boldsymbol{E}^*$ 和 $\boldsymbol{F}^*$ 中元的系数, $\boldsymbol{G}$ 中元的导数是 $(\boldsymbol{E}+\boldsymbol{E}^*)'(\boldsymbol{E}+\boldsymbol{E}^*)$ 的展开式中 $\boldsymbol{E}^*$ 和 $\boldsymbol{F}^*$ 中元的系数. 因此从 $\boldsymbol{A},\boldsymbol{G}$ 到 $\boldsymbol{E},\boldsymbol{F}$ 变换的雅可比行列式是下线性变换的行列式,

$$\mathrm{d}\boldsymbol{A}=(\mathrm{d}\boldsymbol{E})'\boldsymbol{F}\boldsymbol{E}+\boldsymbol{E}'(\mathrm{d}\boldsymbol{F})\boldsymbol{E}+\boldsymbol{E}'\boldsymbol{F}(\mathrm{d}\boldsymbol{E}), \tag{23}$$

$$\mathrm{d}\boldsymbol{G}=(\mathrm{d}\boldsymbol{E})'\boldsymbol{E}+\boldsymbol{E}'(\mathrm{d}\boldsymbol{E}). \tag{24}$$

由于 $\boldsymbol{A}$ 和 $\boldsymbol{G}$ ($\mathrm{d}\boldsymbol{A}$ 和 $\mathrm{d}\boldsymbol{G}$) 是对称的, 上面仅仅用到了函数独立的分量方程.

(23) 式和 (24) 式左乘 $\boldsymbol{E}'^{-1}$, 右乘 $\boldsymbol{E}^{-1}$, 得到

$$\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{A})\boldsymbol{E}^{-1}=\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{E})'\boldsymbol{F}+\mathrm{d}\boldsymbol{F}+\boldsymbol{F}(\mathrm{d}\boldsymbol{E})\boldsymbol{E}^{-1}, \tag{25}$$

$$\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{G})\boldsymbol{E}^{-1}=\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{E})'+(\mathrm{d}\boldsymbol{E})\boldsymbol{E}^{-1}. \tag{26}$$

要记住, 现在是把 (23) 式和 (24) 式当作线性变换考虑, 而不考虑方程是怎么得到的.

令

$$\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{A})\boldsymbol{E}^{-1}=\mathrm{d}\overline{\boldsymbol{A}}, \tag{27}$$

$$\boldsymbol{E}'^{-1}(\mathrm{d}\boldsymbol{G})\boldsymbol{E}^{-1}=\mathrm{d}\overline{\boldsymbol{G}}, \tag{28}$$

$$(\mathrm{d}\boldsymbol{E})\boldsymbol{E}^{-1}=\mathrm{d}\boldsymbol{W}. \tag{29}$$

那么

$$\mathrm{d}\overline{\boldsymbol{A}}=(\mathrm{d}\boldsymbol{W})'\boldsymbol{F}+\mathrm{d}\boldsymbol{F}+\boldsymbol{F}(\mathrm{d}\boldsymbol{W}), \tag{30}$$

$$\mathrm{d}\overline{\boldsymbol{G}}=\mathrm{d}\boldsymbol{W}'+\mathrm{d}\boldsymbol{W}. \tag{31}$$

可以把从 $\mathrm{d}\boldsymbol{E},\mathrm{d}\boldsymbol{F}$ 到 $\mathrm{d}\boldsymbol{A},\mathrm{d}\boldsymbol{G}$ 的线性变换当作从 $\mathrm{d}\boldsymbol{E},\mathrm{d}\boldsymbol{F}$ 到 $\mathrm{d}\boldsymbol{W},\mathrm{d}\boldsymbol{F}$ 的线性变换, 它有行列式 $|\boldsymbol{E}^{-1}|^p=|\boldsymbol{E}|^{-p}$ ($\mathrm{d}\boldsymbol{E}$ 的每行是由 $\boldsymbol{E}^{-1}$ 变换的), 接下来是从 $\mathrm{d}\boldsymbol{W},\mathrm{d}\boldsymbol{F}$ 到 $\mathrm{d}\overline{\boldsymbol{A}},\mathrm{d}\overline{\boldsymbol{G}}$ 的线性变换, 还有从 $\mathrm{d}\overline{\boldsymbol{A}},\mathrm{d}\overline{\boldsymbol{G}}$ 到 $\mathrm{d}\boldsymbol{A}=\boldsymbol{E}'(\mathrm{d}\overline{\boldsymbol{A}})\boldsymbol{E},\mathrm{d}\boldsymbol{G}=\boldsymbol{E}'(\mathrm{d}\overline{\boldsymbol{G}})\boldsymbol{E}$ 的线性变换, 有行列式 $|\boldsymbol{E}|^{p+1}\cdot|\boldsymbol{E}|^{p+1}$(7.3.3 节); 从 $\mathrm{d}\boldsymbol{E},\mathrm{d}\boldsymbol{F}$ 到 $\mathrm{d}\boldsymbol{A},\mathrm{d}\boldsymbol{G}$ 的线性变换的行列式是三个分量变换的行列式之积. 变换 (30), (31) 写成分量形式为

$$\begin{aligned}
\mathrm{d}\bar{a}_{ii}&=\mathrm{d}f_i+2f_i\mathrm{d}w_{ii},\\
\mathrm{d}\bar{a}_{ij}&=f_j\mathrm{d}w_{ji}+f_i\mathrm{d}w_{ij},\quad i<j,\\
\mathrm{d}\bar{g}_{ii}&=2\mathrm{d}w_{ii},\\
\mathrm{d}\bar{g}_{ij}&=\mathrm{d}w_{ji}+\mathrm{d}w_{ij},\quad i<j.
\end{aligned} \tag{32}$$

行列式是

$$\begin{array}{c|cccc|}
 & \mathrm{d}f_i & \mathrm{d}w_{ii} & \mathrm{d}w_{ij}(i<j) & \mathrm{d}w_{ij}(i>j)\\
\mathrm{d}\bar{a}_{ii} & \boldsymbol{I} & 2\boldsymbol{F} & \boldsymbol{0} & \boldsymbol{0}\\
\mathrm{d}\bar{g}_{ii} & \boldsymbol{0} & 2\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0}\\
\mathrm{d}\bar{a}_{ij}(i<j) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M} & \boldsymbol{N}\\
\mathrm{d}\bar{g}_{ij}(i<j) & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{I}
\end{array} \tag{33}$$

$$= \begin{vmatrix} I & 2F \\ 0 & 2I \end{vmatrix} \cdot \begin{vmatrix} M & N \\ I & I \end{vmatrix} = 2^p |M - N|,$$

其中

$$M = \begin{array}{c} \\ da_{12} \\ \vdots \\ da_{1p} \\ da_{23} \\ \vdots \\ da_{2p} \\ \vdots \\ da_{p-1,p} \end{array} \begin{array}{c} \begin{array}{ccccccc} dw_{12} & \cdots & dw_{1p} & dw_{23} & \cdots & dw_{2p} & \cdots & dw_{p-1,p} \end{array} \\ \left[\begin{array}{ccc|ccc|c|c} f_1 & \cdots & 0 & 0 & \cdots & 0 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & f_1 & 0 & \cdots & 0 & & 0 \\ \hline 0 & \cdots & 0 & f_2 & \cdots & 0 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & f_2 & & 0 \\ \hline \vdots & & & & & & & \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & & f_{p-1} \end{array}\right] \end{array} \tag{34}$$

$$N = \begin{array}{c} \\ da_{12} \\ \vdots \\ da_{1p} \\ da_{23} \\ \vdots \\ da_{2p} \\ \vdots \\ da_{p-1,p} \end{array} \begin{array}{c} \begin{array}{ccccccc} dw_{12} & \cdots & dw_{p1} & dw_{32} & \cdots & dw_{p2} & \cdots & dw_{p,p-1} \end{array} \\ \left[\begin{array}{ccc|ccc|c|c} f_2 & \cdots & 0 & 0 & \cdots & 0 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & f_p & 0 & \cdots & 0 & & 0 \\ \hline 0 & \cdots & 0 & f_3 & \cdots & 0 & & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & f_p & & 0 \\ \hline \vdots & & & & & & & \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & & f_p \end{array}\right] \end{array} \tag{35}$$

则

$$|M - N| = \prod_{i<j} (f_i - f_j). \tag{36}$$

线性变换 (23), (24) 的行列式是

$$|E|^{-p} |E|^{p+1} |E|^{p+1} 2^p \prod_{i<j} (f_i - f_j) = 2^p |E|^{p+2} \prod_{i<j} (f_i - f_j). \tag{37}$$

定理 13.2.1 变换 (19) 的雅可比行列式是 (37) 的绝对值.

13.2.3 矩阵 E 和根的联合分布

A 和 B 的联合密度是

$$w(A|I, m) w(B|I, n) = C_1 |A|^{\frac{1}{2}(m-p-1)} |B|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\mathrm{tr}(A+B)}, \tag{38}$$

其中

$$C_1 = \left[2^{\frac{1}{2}p(n+m)} \Gamma_p\left(\frac{1}{2}n\right) \Gamma_p\left(\frac{1}{2}m\right) \right]^{-1}. \tag{39}$$

因此 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 的联合密度是

$$C_1|\boldsymbol{E}'\boldsymbol{F}\boldsymbol{E}|^{\frac{1}{2}(m-p-1)}|\boldsymbol{E}'(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{E}|^{\frac{1}{2}(n-p-1)}\cdot \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{E}'\boldsymbol{E}}2^p|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}(p+2)}\prod_{i<j}(f_i-f_j). \tag{40}$$

因为 $|\boldsymbol{E}'\boldsymbol{F}\boldsymbol{E}|=|\boldsymbol{E}'|\cdot|\boldsymbol{F}|\cdot|\boldsymbol{E}|=|\boldsymbol{F}|\cdot|\boldsymbol{E}'\boldsymbol{E}|=\prod_{i=1}^p f_i|\boldsymbol{E}'\boldsymbol{E}|$ 和 $|\boldsymbol{E}'(\boldsymbol{I}-\boldsymbol{F})\boldsymbol{E}|=|\boldsymbol{I}-\boldsymbol{F}|\cdot|\boldsymbol{E}'\boldsymbol{E}|=\prod_{i=1}^p(1-f_i)|\boldsymbol{E}'\boldsymbol{E}|$, 所以 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 的密度是

$$2^pC_1|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}(m+n-p)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{E}'\boldsymbol{E}}\prod_{i=1}^p f_i^{\frac{1}{2}(m-p-1)}\prod_{i=1}^p(1-f_i)^{\frac{1}{2}(n-p-1)}\prod_{i<j}(f_i-f_j). \tag{41}$$

显然, $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是统计独立的, 因为密度可以分解成 $\boldsymbol{E}$ 的函数和 $\boldsymbol{F}$ 的函数. 为了确定边缘密度我们只需要找到两个正规化常数 (积是 2^pC_1).

我们求

$$2^p\int|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}(m+n-p)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{E}'\boldsymbol{E}}\mathrm{d}\boldsymbol{E}, \tag{42}$$

其中积分是 $0<e_{i1}<\infty,-\infty<e_{ij}<\infty,j\neq 1$. 如果我们令 $-\infty<e_{i1}<\infty$, 再用 2^{-p} 乘则 (42) 式的值不变. 因此 (42) 式是

$$(2\pi)^{\frac{1}{2}p^2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}(m+n-p)}\left[\frac{1}{(2\pi)^{\frac{1}{2}p^2}}\exp\left(-\frac{1}{2}\sum_{i,j}e_{ij}^2\right)\right]\prod \mathrm{d}e_{ij}. \tag{43}$$

除了常数 $(2\pi)^{\frac{1}{2}p^2}$, (43) 式是当 e_{ij} 有方括号中函数的密度时 $|\boldsymbol{E}'\boldsymbol{E}|$ 的 $\frac{1}{2}(m+n-p)$ 次幂的期望值的定义. 这个期望值是当 $|\boldsymbol{E}'\boldsymbol{E}|$ 有分布 $W(\boldsymbol{I},p)$ 时 $|\boldsymbol{E}'\boldsymbol{E}|$ 的广义方差的 $\frac{1}{2}(m+n-p)$ 阶矩. (见 7.5 节.) 因此 (43) 式是

$$(2\pi)^{\frac{1}{2}p^2}\frac{\boldsymbol{\Gamma}_p[\frac{1}{2}(m+n)]}{\boldsymbol{\Gamma}_p(\frac{1}{2}p)}2^{\frac{1}{2}p(n+m-p)}. \tag{44}$$

因此 $\boldsymbol{E}$ 的密度是

$$\frac{\boldsymbol{\Gamma}_p(\frac{1}{2}p)}{2^{\frac{1}{2}p(m+n-2)}\pi^{\frac{1}{2}p^2}\boldsymbol{\Gamma}_p[\frac{1}{2}(m+n)]}|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}(m+n-p)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{E}'\boldsymbol{E}}. \tag{45}$$

f_i 的密度是 (41) 式按 (45) 式划分, 即对 $0\leqslant f_p\leqslant\cdots\leqslant f_1\leqslant 1$, f_i 的密度是

$$C_2\prod_{i=1}^p f_i^{\frac{1}{2}(m-p-1)}\prod_{i=1}^p(1-f_i)^{\frac{1}{2}(n-p-1)}\prod_{i<j}(f_i-f_j), \tag{46}$$

其中

$$C_2=\frac{\pi^{\frac{1}{2}p^2}\boldsymbol{\Gamma}_p[\frac{1}{2}(m+n)]}{\boldsymbol{\Gamma}_p(\frac{1}{2}n)\boldsymbol{\Gamma}_p(\frac{1}{2}m)\boldsymbol{\Gamma}_p(\frac{1}{2}p)}. \tag{47}$$

令

$$f_i=\frac{l_i}{l_i+1}, \tag{48}$$

就可从 (46) 式得到 l_i 的密度; 我们有

$$\frac{\mathrm{d}f_i}{\mathrm{d}l_i}=\frac{1}{(l_i+1)^2},$$
$$f_i-f_j=\frac{l_i-l_j}{(l_i+1)(l_j+1)},$$
$$1-f_i=\frac{1}{l_i+1}. \tag{49}$$

因此对 $0\leqslant l_p\leqslant\cdots\leqslant l_1$, l_i 的密度是

$$C_2\prod_{i=1}^{p}l_i^{\frac{1}{2}(m-p-1)}\prod_{i=1}^{p}(l_i+1)^{-\frac{1}{2}(m+n)}\prod_{i<j}(l_i-l_j). \tag{50}$$

定理 13.2.2 如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是独立地分别服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)$ 分布 $(m\geqslant p,n\geqslant p)$, $|\boldsymbol{A}-l\boldsymbol{B}|=0$ 的根的联合密度是 (50) 式, 其中 C_2 是 (47) 式定义的.

$\boldsymbol{Y}$ 的联合密度可以从 (45) 式得到, 事实上雅可比行列式是 $|\boldsymbol{Y}|^{-2p}$. (见附录的定理 A.4.6.)

13.2.4 $\boldsymbol{A}$ 奇异时的分布

上面的矩阵 $\boldsymbol{A}$ 可以表示成 $\boldsymbol{A}=\boldsymbol{W}_1\boldsymbol{W}_1'$, 其中 $\boldsymbol{W}_1$ $(p\times m)$ 的列是独立地服从分布 $N(0,\boldsymbol{\Sigma})$ 的. 我们现在考虑 $m<p$ 时的情况. 如果我们令 $\boldsymbol{B}+\boldsymbol{W}_1\boldsymbol{W}_1'=\boldsymbol{G}=\boldsymbol{C}\boldsymbol{C}'$, $\boldsymbol{W}_1=\boldsymbol{C}\boldsymbol{U}$, 那么

$$\begin{aligned}0&=|\boldsymbol{A}-f(\boldsymbol{A}+\boldsymbol{B})|=|\boldsymbol{W}_1\boldsymbol{W}_1'-f\boldsymbol{G}|\\&=|\boldsymbol{C}\boldsymbol{U}\boldsymbol{U}'\boldsymbol{C}'-f\boldsymbol{C}\boldsymbol{C}'|=|\boldsymbol{C}|\cdot|\boldsymbol{U}\boldsymbol{U}'-f\boldsymbol{I}_p|\cdot|\boldsymbol{C}|\end{aligned} \tag{51}$$

的根也即

$$|\boldsymbol{U}\boldsymbol{U}'-f\boldsymbol{I}_p|=0 \tag{52}$$

的根. 我们应该证明非零根 $f_1>\cdots>f_m$(这些根是以概率不同的 1) 是

$$|\boldsymbol{U}\boldsymbol{U}'-f\boldsymbol{I}_m|=0 \tag{53}$$

的根. 对 (52) 式的每一个根 $f\neq0$, 均有向量 $\boldsymbol{x}$ 满足

$$(\boldsymbol{U}'\boldsymbol{U}-f\boldsymbol{I}_p)\boldsymbol{x}=\boldsymbol{0}. \tag{54}$$

左乘 $\boldsymbol{U}'$ 给出

$$\begin{aligned}\boldsymbol{0}&=\boldsymbol{U}'(\boldsymbol{U}\boldsymbol{U}'-f\boldsymbol{I}_p)\boldsymbol{x}\\&=(\boldsymbol{U}'\boldsymbol{U}-f\boldsymbol{I}_m)(\boldsymbol{U}'\boldsymbol{x}).\end{aligned} \tag{55}$$

因此 $\boldsymbol{U}'\boldsymbol{x}$ 是 $\boldsymbol{U}'\boldsymbol{U}$ 的特征向量, f 是相应的根.

8.4 节中证明了 $\boldsymbol{U}=\boldsymbol{U}_*'$ 的密度是 (当 $\boldsymbol{I}_p-\boldsymbol{U}\boldsymbol{U}'$ 正定或者 $\boldsymbol{I}_m-\boldsymbol{U}*\boldsymbol{U}_*'$ 正定时)

$$K|\boldsymbol{I}_p-\boldsymbol{U}\boldsymbol{U}'|^{\frac{1}{2}(n-p-1)}=K|\boldsymbol{I}_{p^*}-\boldsymbol{U}*\boldsymbol{U}_*'|^{\frac{1}{2}(n^*-p^*-1)}, \tag{56}$$

其中 $p^*=m,n^*-p^*-1=n-p-1,m^*=p$. 因此 $f_1,\cdots,f_m$ 服从 (46) 式中 p 被 m 取代 m 被 p 取代, n 被 $n+m-p$ 取代的分布, 即

$$\frac{\pi^{\frac{1}{2}m^2}\Gamma_m[\frac{1}{2}(m+n)]}{\Gamma_m(\frac{1}{2}m)\Gamma_m[\frac{1}{2}(m+n-p)]\Gamma_m(\frac{1}{2}p)} \tag{57}$$
$$\cdot\prod_{i=1}^{m}\left[f_i^{\frac{1}{2}(p-m-1)}(1-f_i)^{\frac{1}{2}(n-p-1)}\right]\prod_{i<j}(f_i-f_j).$$

定理 13.2.3 *如果* A *与* W_1W_1' *同分布, 其中* W_1 *的* m *列是相互独立的, 均服从* $N(\mathbf{0},\Sigma)$ *分布,* $m\leqslant p$, B *独立地服从分布* $W(\Sigma,n), n\geqslant p$, *则* $|A-f(A+B)|=0$ *由式 (57) 给出.*

根的分布是独立的, 大约同一时间被 Fisher(1939), Girshick(1939), Hsu(1939a), Mood(1951), Roy(1939) 得到. 13.2.2 节雅可比行列式的发展主要是归功于 Hsu[如 Deemer and Olkin(1951) 所说].

13.3 一个非奇异 Wishart 矩阵的情况

本节我们寻找

$$|A-lI|=0 \tag{1}$$

的根的分布, 其中矩阵 A 有分布 $W(I,n)$. 将会观察到来自分布 $N(\mu,I)$ 的 $n+1$ 个样本的主成分的方差是 $1/n$ 倍的 (1) 式的根. 我们得到下面的有用的定理.

定理 13.3.1 *如果对称矩阵* B *有密度形式* $g(l_1,\cdots,l_p)$, *其中* $l_1>\cdots>l_p$ *是* B *的特征根, 则根的密度是*

$$\frac{\pi^{\frac{1}{2}p^2}g(l_1,\cdots,l_p)\prod_{i<j}(l_i-i_j)}{\Gamma_p(\frac{1}{2}p)}. \tag{2}$$

证明 由附录的定理 A.2.1, 我们得到存在一个正交矩阵 C 满足

$$B=C'LC, \tag{3}$$

其中

$$L=\begin{pmatrix} l_1 & 0 & \cdots & 0\\ 0 & l_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & l_p\end{pmatrix}. \tag{4}$$

如果这些 l 按大小递减次序排序, 并且如果 $c_{i1}\geqslant 0$, 则 (以概率 1) 从 B 到 L 和 C 的变换是唯一的. 令矩阵 C 的坐标是 $c_1,\cdots,c_{p(p-1)/2}$, 令变换的雅可比行列式是 $f(L,C)$. 则 L 和 C 的联合密度是 $g(l_1,\cdots,l_p)f(L,C)$. 为了证明定理, 我们必须证明

$$\int\cdots\int f(L,C)\mathrm{d}c_1\cdots\mathrm{d}c_{p(p-1)/2}=\frac{\pi^{\frac{1}{2}p^2}\prod_{i<j}(l_i-l_j)}{\Gamma_p(\frac{1}{2}p)}. \tag{5}$$

我们现在在一种特殊的情况下证明, 即 $B=UU'$, U $(p\times m, m\geqslant p)$ 有密度

$$\pi^{-\frac{1}{2}mp}\frac{\Gamma_p[\frac{1}{2}(m+n)]}{\Gamma_p(\frac{1}{2}n)}|\boldsymbol{I}-\boldsymbol{U}\boldsymbol{U}'|^{\frac{1}{2}(n-p-1)}. \tag{6}$$

然后由下面将叙述的引理 13.3.1, 得到 $\boldsymbol{B}$ 有密度

$$\begin{aligned}&\frac{\Gamma_p[\frac{1}{2}(m+n)]}{\Gamma_p(\frac{1}{2}m)\Gamma_p(\frac{1}{2}n)}|\boldsymbol{I}-\boldsymbol{B}|^{\frac{1}{2}(n-p-1)}|\boldsymbol{B}|^{\frac{1}{2}(m-p-1)} \\ &=\frac{\Gamma_p[\frac{1}{2}(m+n)]}{\Gamma_p(\frac{1}{2}m)\Gamma_p(\frac{1}{2}n)}\prod_{i=1}^{p}(1-l_i)^{\frac{1}{2}(n-p-1)}\prod_{i=1}^{p}l_i^{\frac{1}{2}(m-p-1)} \\ &=g^*(l_1,\cdots,l_p).\end{aligned} \tag{7}$$

$\boldsymbol{L}$ 和 $\boldsymbol{C}$ 的联合密度是 $f(\boldsymbol{L},\boldsymbol{C})g^*(l_1,\cdots,l_p)$. 在前面的章节我们证明了 $\boldsymbol{L}$ 的边缘密度是式 (50). 因此

$$\begin{aligned}\int\cdots\int g^*(l_1,\cdots,l_p)f(\boldsymbol{L},\boldsymbol{C})\mathrm{d}\boldsymbol{C}&=g^*(l_1,\cdots,l_p)\int\cdots\int f(\boldsymbol{L},\boldsymbol{C})\mathrm{d}\boldsymbol{C} \\ &=\frac{\pi^{\frac{1}{2}p^2}\prod(l_i-l_j)}{\Gamma_p(\frac{1}{2}p)}g^*(l_1,\cdots,l_p).\end{aligned} \tag{8}$$

这个证明了 (5) 式, 因此证明了定理. ■

(7) 式上面的说明基于下面的引理.

引理 13.3.1 *如果 $\boldsymbol{Y}$ $(p\times m)$ 的密度是 $f(\boldsymbol{Y}\boldsymbol{Y}')$, 则 $\boldsymbol{B}=\boldsymbol{Y}\boldsymbol{Y}'$ 的密度是*

$$\frac{|\boldsymbol{B}|^{\frac{1}{2}(m-p-1)}f(\boldsymbol{B})\pi^{\frac{1}{2}pm}}{\Gamma_p(\frac{1}{2}m)}. \tag{9}$$

这个引理的证明, 类似于定理 13.3.1 的证明, 也是依靠一个特殊的例子; 令 $f(\boldsymbol{Y}\boldsymbol{Y}')=(2\pi)^{-\frac{1}{2}pm}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{Y}\boldsymbol{Y}'}$, 则 (9) 式是 $w(\boldsymbol{B}|\boldsymbol{I},m)$.

现在我们求 (1) 的根的密度. $\boldsymbol{A}$ 的密度是

$$\frac{|\boldsymbol{A}|^{\frac{1}{2}(n-p-1)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{A}}}{2^{\frac{1}{2}pn}\Gamma_p(\frac{1}{2}n)}=\frac{\prod_{i=1}^{p}l_i^{\frac{1}{2}(n-p-1)}\exp(-\frac{1}{2}\sum_{i=1}^{p}l_i)}{2^{\frac{1}{2}pn}\Gamma_p(\frac{1}{2}n)}. \tag{10}$$

因此由定理我们得到了 $\boldsymbol{A}$ 的根的密度

$$\frac{\pi^{\frac{1}{2}p^2}\prod_{i=1}^{p}l_i^{\frac{1}{2}(n-p-1)}\exp(-\frac{1}{2}\sum_{i=1}^{p}l_i)\prod_{i<j}(l_i-l_j)}{2^{\frac{1}{2}pn}\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}p)}. \tag{11}$$

定理 13.3.2 *如果 $\boldsymbol{A}$ $(p\times p)$ 有分布 $\boldsymbol{W}(\boldsymbol{I},n)$, 那么特征根 $(l_1\geqslant l_2\geqslant\cdots\geqslant l_p\geqslant 0)$ 在密度不为 0 的区域有密度 (11).*

推论 13.3.1 *令 $v_1\geqslant\cdots\geqslant v_p$ 是来自总体 $N(\boldsymbol{\mu},\sigma^2\boldsymbol{I})$ 的 $\boldsymbol{N}=n+1$ 个样本的样本主成分的样本方差. 则 $(n/\sigma^2)v_i$ 的分布有密度 (11).*

$\boldsymbol{A}$ 的特征向量是以概率 1 由

$$(\boldsymbol{A}-l\boldsymbol{I})\boldsymbol{y}=\boldsymbol{0},\quad \boldsymbol{y}'\boldsymbol{y}=1 \tag{12}$$

唯一定义的 (除了 -1 倍数外). 因为根不同的概率是 1. 令有 $y_{1i}\geqslant 0$ 的向量是

$$Y=(y_1,\cdots,y_p). \tag{13}$$

则有

$$AY=YL. \tag{14}$$

从 11.2 节我们知道

$$Y'Y=I. \tag{15}$$

将 (14) 式右端乘 $Y^{-1}=Y'$, 得到

$$A=YLY'. \tag{16}$$

因此 $Y'=C$, C 在前面已定义.

现在我们考虑 L 和 C 的联合分布. 矩阵 A 与

$$A=\sum_{\alpha=1}^{n}X_\alpha X_\alpha' \tag{17}$$

同分布, 其中 X_α 是独立地分布的, 均服从 $N(0,I)$. 令

$$X_\alpha^*=QX_\alpha, \tag{18}$$

其中 Q 是任意正交矩阵. 则 X_α^* 是独立地服从 $N(0,I)$ 分布的,

$$A^*=\sum_{\alpha=1}^{n}X_\alpha^*X_\alpha^{*'}=QAQ' \tag{19}$$

服从 $W(I,n)$ 分布. A^* 的根是 A 的根; 因此如果我们要求 $c_{i1}^{**}\geqslant 0$ 那么

$$A^*=C^{**'}LC^{**}, \tag{20}$$

$$C^{**'}C^{**}=I \tag{21}$$

定义了 C^{**}. 令

$$C^*=CQ'. \tag{22}$$

令

$$J(C^*)=\begin{pmatrix}\frac{c_{11}^*}{|c_{11}^*|} & 0 & \cdots & 0\\ 0 & \frac{c_{21}^*}{|c_{21}^*|} & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \frac{c_{p1}^*}{|c_{p1}^*|}\end{pmatrix}, \tag{23}$$

如果 $c_{i1}^*=0$, 那么 $c_{i1}^*/|c_{i1}^*|=1$. 因此 $J(C^*)$ 是一个对角矩阵; 如果 $c_{i1}^*\geqslant 0$, 那么第 i 个对角元是 1, 如果 $c_{i1}^*<0$, 那么第 i 个对角元是 -1. 因此

$$C^{**}=J(C^*)C^*=J(CQ')CQ'. \tag{24}$$

C^{**} 的分布同 C 的分布相同. 我们现在将证明这个事实定义了 C 的分布.

定义 13.3.1 *如果 p 阶随机正交矩阵 E 有分布使得 EQ' 对每个正交的 Q 有相同的分布, 则称 E 有 Haar 不变分布 (或者正规化度量).*

定义是可能的, 因为已经证明了仅存在一个分布满足要求的不变性质 [Halmos(1950)]. 已经证明了这个分布在左端乘一个正交矩阵下是唯一的不变量 (即

$\boldsymbol{QE}$ 的分布同 $\boldsymbol{E}$ 相同). 由此得到 $\boldsymbol{E}$ 有 $e_{i1} \geqslant 0$ 的概率是 $1/2^p$. 这些将在下面证明. 令 $\boldsymbol{J}_1, \cdots, \boldsymbol{J}_{2^p}$ 是 2^p 个对角矩阵, 对角元素是 $+1$ 和 -1. 由于 $\boldsymbol{J}_i\boldsymbol{E}$ 的分布同 $\boldsymbol{E}$ 的分布相同, $e_{i1} \geqslant 0$ 的概率同 $\boldsymbol{J}_i\boldsymbol{E}$ 的第一列元素非负的概率相同. 这些事件对 $i = 1, \cdots, 2^p$ 是相互排斥的和穷举的 (除了 0 元素外, 它有概率 0), 因此任意一个的概率是 $1/2^p$.

在空间的这个部分上, 给定 $e_{i1} \geqslant 0$ 时 $\boldsymbol{E}$ 的条件分布是 2^p 倍的 Haar 不变分布. 我们称它为条件 Haar 不变分布.

引理 13.3.2　*如果正交矩阵 $\boldsymbol{E}$ 有分布使得 $e_{i1} \geqslant 0$, $\boldsymbol{E}^{**} = \boldsymbol{J}(\boldsymbol{EQ}')\boldsymbol{EQ}'$ 对每个正交的 $\boldsymbol{Q}$ 有相同的分布, 则 $\boldsymbol{E}$ 有条件 Haar 不变分布.*

证明　令正交矩阵的空间 V 分成子空间 $V_1, \cdots, V_{2^p}$, 使得 $\boldsymbol{J}_iV_i = V_1$, 其中 $\boldsymbol{J}_1 = \boldsymbol{I}$, V_1 是当 $e_{i1} \geqslant 0$ 时的集合. 令 μ_1 是由引理中假设的 $\boldsymbol{E}$ 的分布定义的 V_1 的测度. V_i 中的集合 (可测的)W 的测度 $\mu(W)$ 被定义成 $(1/2^p)\mu_1(\boldsymbol{J}_iW)$. 现在我们想要证明 μ 是 Haar 不变测度. 令 W 是 V_1 中任意 (可测的) 集合. 引理假设 $2^p\mu(W) = \mu_1(W) = \Pr\{\boldsymbol{E} \in W\} = \Pr\{\boldsymbol{E}^{**} \in W\} = \sum \mu_1(\boldsymbol{J}_i[W\boldsymbol{Q}' \cap V_i]) = 2^p\mu(W\boldsymbol{Q}')$. 如果 U 是 V 中的任意 (可测的) 集合, 则 $U = \cup_{j=1}^{2^p}(U \cap V_j)$. 因为 $\mu(U \cap V_j) = (1/2^p)\mu_1[\boldsymbol{J}_j(U \cap V_j)]$, 由上面所述知, 这个是 $\mu[(U \cap V_j)\boldsymbol{Q}']$. 因此 $\mu(U) = \mu(U\boldsymbol{Q}')$. 因此 μ 是不变的, μ_1 是条件不变分布. ■

从引理我们看出矩阵 $\boldsymbol{C}$ 有条件 Haar 不变分布. 因为它和 $\boldsymbol{C}$ 基于 $\boldsymbol{L}$ 的条件分布相同, $\boldsymbol{C}$ 和 $\boldsymbol{L}$ 是独立的.

定理 13.3.3　*如果 $\boldsymbol{C} = \boldsymbol{Y}'$, 其中 $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_p)$ 是 $y_{i1} \geqslant 0$ 时的 $\boldsymbol{A}$ 的正规化特征向量, 其中 $\boldsymbol{A}$ 服从 $W(\boldsymbol{I}, n)$ 分布, 则 $\boldsymbol{C}$ 有条件 Haar 不变分布, $\boldsymbol{C}$ 和特征根是独立分布.*

从先前的工作中我们可以推广定理 13.3.1.

定理 13.3.4　*如果对称矩阵 $\boldsymbol{B}$ 有密度形式 $g(l_1, \cdots, l_p)$, 其中 $l_1 > \cdots > l_p$ 是 $\boldsymbol{B}$ 的特征根, 则根的联合密度是 (2) 式, 正规化特征向量 $\boldsymbol{Y}(y_{1i} \geqslant 0)$ 的矩阵独立地服从条件 Haar 不变分布.*

证明　$\boldsymbol{QBQ}'$ 的密度, 同 $\boldsymbol{B}$(根是不变的) 的密度相同, 其中 $\boldsymbol{QQ}' = \boldsymbol{I}$, 因此 $\boldsymbol{J}(\boldsymbol{Y}'\boldsymbol{Q}')\boldsymbol{Y}'\boldsymbol{Q}'$ 的分布同 $\boldsymbol{Y}'$ 的分布相同. 则定理 13.3.4 由引理 13.3.2 得证. ■

我们给出这个定理的一个应用, 在 $\boldsymbol{B} = \boldsymbol{B}'$ 是正态分布 (函数独立的) 情况下, 其中 $\boldsymbol{B}$ 的分量相互独立, 有均值 0, 方差 $\mathrm{E}(b_{ii}^2) = 1$ 和 $\mathrm{E}(b_{ij}^2) = \frac{1}{2}(i < j)$.

定理 13.3.5　*令 $\boldsymbol{B} = \boldsymbol{B}'$ 有密度*

$$\pi^{-p(p+1)/4}2^{-\frac{1}{2}p}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\boldsymbol{B}^2}. \tag{25}$$

则 $\boldsymbol{B}$ 的特征根 $l_1 > \cdots > l_p$ 有密度

$$2^{-\frac{1}{2}p}\pi^{p(p-1)/4}\Gamma_p^{-1}\left(\frac{1}{2}p\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{p}l_i^2\right)\prod_{i<j}(l_i - l_j), \tag{26}$$

正规化特征向量 ($y_{1i} \geqslant 0$) 的矩阵 $\boldsymbol{Y}$ 独立地服从条件 Haar 不变分布.

证明 因为 $\boldsymbol{B}^2$ 的特征根是 $l_1^2, \cdots, l_p^2$, $\operatorname{tr}\boldsymbol{B}^2 = \sum l_i^2$, 定理直接得证. ■

推论 13.3.2 令 $n\boldsymbol{S}$ 是服从 $W(\boldsymbol{I}, n)$ 分布的, 由 $\boldsymbol{S}=\boldsymbol{C}'\boldsymbol{L}\boldsymbol{C}$ 和 $\boldsymbol{C}'\boldsymbol{C}=\boldsymbol{I}(l_1 > \cdots > l_p)$ 定义对角矩阵 $\boldsymbol{L}$ 和 $\boldsymbol{B}$, $c_{i1} \geqslant 0, i = 1, \cdots, p$. 则 $\sqrt{n}(\boldsymbol{L}-\boldsymbol{I}) = \boldsymbol{D}$ 对角的极限分布的密度是 (26) 式, l_i 被 d_i 取代, 矩阵 $\boldsymbol{C}$ 根据条件 Haar 测度是独立分布的.

证明 $\sqrt{n}(\boldsymbol{S}-\boldsymbol{I})$ 的极限分布的密度是 (25) 式, $\boldsymbol{D}$ 的对角元是 $\sqrt{n}(\boldsymbol{S}-\boldsymbol{I})$ 的特征根, $\boldsymbol{C}'$ 的列是特征向量. ■

13.4 典型相关

在 12.3 节中证明了样本典型相关是

$$|\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} - f\boldsymbol{A}_{11}| = 0 \tag{1}$$

的根的平方根, 其中

$$\boldsymbol{A}_{ij} = \sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha^{(i)} - \bar{\boldsymbol{X}}^{(i)})(\boldsymbol{X}_\alpha^{(j)} - \bar{\boldsymbol{X}}^{(j)})', \qquad i,j = 1,2, \tag{2}$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \end{pmatrix} \tag{3}$$

的分布是 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 其中

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{4}$$

从 3.3 节我们知道 $\boldsymbol{A}_{ij}$ 的分布同

$$\boldsymbol{A}_{ij} = \sum_{\alpha=1}^{n} \boldsymbol{Y}_\alpha^{(i)}\boldsymbol{Y}_\alpha^{(j)\prime}, \quad i,j = 1,2 \tag{5}$$

的分布相同, 其中 $n = N-1$,

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}^{(1)} \\ \boldsymbol{Y}^{(2)} \end{pmatrix} \tag{6}$$

服从 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ 分布. 我们假设 $\boldsymbol{Y}^{(1)}$ 的维数为 p_1, 它不大于 $\boldsymbol{Y}^{(2)}$ 的维数 p_2. 则 (1) 式有 p_1 个非零的根, 设为

$$f_1 > f_2 > \cdots > f_{p_1}. \tag{7}$$

现在我们求得 $\{f_i\}$ 的分布, 令

$$\boldsymbol{\Sigma}_{12} = \boldsymbol{0}. \tag{8}$$

现在假设 $\{\boldsymbol{Y}_\alpha^{(2)}\}$ 需要确定. $\boldsymbol{A}_{22}$ 是固定的,

$$\boldsymbol{B} = \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \tag{9}$$

是 $\boldsymbol{Y}^{(1)}$ 对 $\boldsymbol{Y}^{(2)}$ 的回归系数矩阵. 从 4.3 节我们知道

$$A_{11.2}=\sum_{\alpha=1}^{n}(Y_\alpha^{(1)}-BY_\alpha^{(2)})(Y_\alpha^{(1)}-BY_\alpha^{(2)})'=A_{11}-BA_{22}B' \tag{10}$$
$$=A_{11}-A_{12}A_{22}^{-1}A_{21}$$

和

$$Q=BA_{22}B'=A_{12}A_{22}^{-1}A_{21} \tag{11}$$

($\boldsymbol{\beta}=0$) 各自服从 $W(\Sigma_{11},n-p_2)$ 和 $W(\Sigma_{11},p_2)$ 分布且独立. 用 Q 的方法, 方程 (1) 定义 f 是

$$|Q-f(A_{11.2}+Q)|=0. \tag{12}$$

$f_i(i=1,\cdots,p_1)$ 的分布, 是 (12) 的非零根的分布, 密度由下式给出 (见 13.2 节),

$$\pi^{\frac{1}{2}p_1^2}\frac{\Gamma_{p_1}(\frac{1}{2}n)}{\Gamma_{p_1}\left[\frac{1}{2}(n-p_2)\right]\Gamma_p(\frac{1}{2}p_1)\Gamma_p(\frac{1}{2}p_2)}$$
$$\cdot\prod_{i=1}^{p_1}\{f_i^{\frac{1}{2}(p_2-p_1-1)}(1-f_i)^{\frac{1}{2}(N-p_2-p_1-2)}\}\prod_{i<j}^{p_1}(f_i-f_j). \tag{13}$$

因为条件密度 (13) 不取决于 $Y^{(2)}$, 所以 (13) 是 $X_\alpha^{(1)}$ 和 $X_\alpha^{(2)}(\alpha=1,\cdots,N)$ 的样本典型相关系数的平方的无条件密度. 当 $X^{(2)}$ 是实际固定的变量向量或者有任意分布时, 密度 (13) 同样成立, 只要 $X^{(1)}$ 和 $X^{(2)}$ 是独立分布的, $X^{(1)}$ 有多元正态分布.

在 $p_1=1,p_2=p-1$ 的特殊情况下, (13) 式简化为

$$\frac{\Gamma\left[\frac{1}{2}(N-1)\right]}{\Gamma\left[\frac{1}{2}(N-p)\right]\Gamma\left[\frac{1}{2}(p-1)\right]}f^{\frac{1}{2}(p-3)}(1-f)^{\frac{1}{2}(N-p-2)}, \tag{14}$$

这是 $X^{(1)}(p_1=1)$ 和 $X^{(2)}(p_2=p-1)$ 之间的样本多重相关系数的平方的密度.

13.5 有一个 Wishart 矩阵情况下的渐近分布

13.5.1 所有总体的根都不同

13.3 节我们求得当 nS 服从 $W(I,n)$ 分布时对角矩阵 L 的密度和正交矩阵 B 的密度, 其中 B 由 $S=BLB'$ 定义 $l_1\geqslant\cdots\geqslant l_p$, $b_{1i}\geqslant 0,i=1,\cdots,p$. 本节我们求当 nS 服从 $W(\Sigma,n)$ 分布时, L 和 B 的渐近分布, 并且 Σ 的特征根是不相同的. (推论 13.3.2 给出了当 $\Sigma=I$ 时的渐近分布.)

定理 13.5.1 *假设 nS 有分布 $W(\Sigma,n)$. 对角矩阵 Λ,L, 和正交矩阵 $\boldsymbol{\beta},B$ 由*

$$\Sigma=\boldsymbol{\beta}\Lambda\boldsymbol{\beta}',\quad S=BLB' \tag{1}$$

定义, 其中 $\lambda_1>\lambda_2>\cdots>\lambda_p,l_1\geqslant l_2\geqslant\cdots\geqslant l_p,\beta_{1i}\geqslant 0,b_{1i}\geqslant 0,i=1,\cdots,p$. 定义 $G=\sqrt{n}(B-\boldsymbol{\beta})$ 和对角矩阵 $D=\sqrt{n}(L-\Lambda)$. 则当 D 和 G 是独立时, D 和 G 的极限分布是正态的, D 的对角元是独立的. 对角元 d_i 有极限分布 $N(0,2\lambda_i^2)$. 在 $G=(g_1,\cdots,g_p)$ 的极限分布中 g_i 的协方差阵是

$$\mathrm{ACov}(g_i) = \sum_{\substack{k=1\\k\neq i}}^{p} \frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\boldsymbol{\beta}_k\boldsymbol{\beta}_k', \tag{2}$$

其中 $\boldsymbol{\beta}=(\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_p)$. 在极限分布中 g_i 和 g_j 的协方差阵是

$$\mathrm{ACov}(g_i,g_j) = -\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_i', \quad i\neq j. \tag{3}$$

证明 矩阵 $n\boldsymbol{T}=n\boldsymbol{\beta}'\boldsymbol{S}\boldsymbol{\beta}$ 是服从 $W(\boldsymbol{\Lambda},n)$ 分布的. 令

$$\boldsymbol{T}=\boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}', \tag{4}$$

其中 $\boldsymbol{Y}$ 是正交的. 为了证明 (4) 唯一地确定 $\boldsymbol{Y}$, 我们需要 $y_{ii}\geqslant 0$. 令 $\sqrt{n}(\boldsymbol{T}-\boldsymbol{\Lambda})=\boldsymbol{U}$ 和 $\sqrt{n}(\boldsymbol{Y}-\boldsymbol{I})=\boldsymbol{W}$. 则 (4) 式可以写成

$$\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{U}=\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}\right)\left(\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{D}\right)\left(\boldsymbol{I}+\frac{1}{\sqrt{n}}\boldsymbol{W}\right)', \tag{5}$$

它等价于

$$\boldsymbol{U}=\boldsymbol{W}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{\Lambda}\boldsymbol{W}'+\frac{1}{\sqrt{n}}(\boldsymbol{W}\boldsymbol{D}+\boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}'+\boldsymbol{D}\boldsymbol{W}')+\frac{1}{n}\boldsymbol{W}\boldsymbol{D}\boldsymbol{W}'. \tag{6}$$

从 $\boldsymbol{I}=\boldsymbol{Y}\boldsymbol{Y}'=[\boldsymbol{I}+(1/\sqrt{n})\boldsymbol{W}][\boldsymbol{I}+(1/\sqrt{n})\boldsymbol{W}']$, 我们有

$$\boldsymbol{0}=\boldsymbol{W}+\boldsymbol{W}'+\frac{1}{\sqrt{n}}\boldsymbol{W}\boldsymbol{W}'. \tag{7}$$

我们稍后继续探索和证明这个方法. 如果我们忽略 (6) 式和 (7) 式中的 $1/\sqrt{n}$ 阶和 $1/n$ 阶的项, 就得到了

$$\boldsymbol{U}=\boldsymbol{W}\boldsymbol{\Lambda}+\boldsymbol{D}+\boldsymbol{\Lambda}\boldsymbol{W}', \tag{8}$$

$$\boldsymbol{0}=\boldsymbol{W}+\boldsymbol{W}'. \tag{9}$$

我们从 (9) 式将 $\boldsymbol{W}'=-\boldsymbol{W}$ 代入 (8) 式, 将结果写成分量的形式, 得到 $w_{ii}=0$,

$$d_i=u_{ii},\quad i=1,\cdots,p, \tag{10}$$

$$w_{ij}=\frac{u_{ij}}{\lambda_j-\lambda_i},\quad i\neq j,\quad i,j=1,\cdots,p. \tag{11}$$

(注意有 $w_{ij}=-w_{ji}$.) 从定理 3.4.4 我们知道, $\boldsymbol{U}$ 的极限正态分布中的函数独立的元的是统计独立的, 有均值 0, 方差 $\mathrm{Var}(u_{ii})=2\lambda_i^2$, $\mathrm{Var}(u_{ij})=\lambda_i\lambda_j, i\neq j$. 则 $\boldsymbol{D}$ 和 $\boldsymbol{W}$ 的极限分布是正态的, $d_1,\cdots,d_p,w_{12},w_{13},\cdots,w_{p-1,p}$ 是独立的, 有均值 0 和方差 $\mathrm{Var}(d_i)=2\lambda_i^2, i=1,\cdots,p$, $\mathrm{Var}(w_{ij})=\lambda_i\lambda_j/(\lambda_i-\lambda_j)^2, j=i+1,\cdots,p, i=1,\cdots,p-1$. $\boldsymbol{B}$ 的每列是相应的 $\pm\boldsymbol{\beta}\boldsymbol{Y}$ 的列; 因为 $\boldsymbol{Y}\xrightarrow{p}\boldsymbol{I}$, 我们有 $\boldsymbol{\beta}\boldsymbol{Y}\xrightarrow{p}\boldsymbol{\beta}$, $\boldsymbol{B}$ 的每列以任意高的概率等同于 $\boldsymbol{\beta}\boldsymbol{Y}$ 的相应的列. 则 $G=\sqrt{n}(\boldsymbol{B}-\boldsymbol{\beta})$ 有极限分布 $\boldsymbol{\beta}\sqrt{n}(Y-\boldsymbol{I})=\boldsymbol{\beta}\boldsymbol{W}$. 渐近方差和协方差随后讨论.

现在我们证明 $\boldsymbol{D}$ 和 $\boldsymbol{W}$ 的极限分布. 方程 $\boldsymbol{T}=\boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}'$ 和 $\boldsymbol{I}=\boldsymbol{Y}\boldsymbol{Y}'$, 条件 $l_1\geqslant\cdots>l_p, y_{ii}>0, i=1,\cdots,p$, 定义一个从 $\boldsymbol{T}$ 到 $\boldsymbol{Y},\boldsymbol{L}$ 的一一变换, 除了 0 测集以外. 从 $\boldsymbol{Y},\boldsymbol{L}$ 到 $\boldsymbol{T}$ 的变换是连续可微的. 其逆在 $\boldsymbol{Y}=\boldsymbol{I}$ 和 $\boldsymbol{L}=\boldsymbol{\Lambda}$ 的邻域是连续可微的, 因为方程 (8) 和 (9) 可以有唯一解. 因此 $\boldsymbol{Y},\boldsymbol{L}$ 作为 $\boldsymbol{T}$ 的函数满足定理 4.2.3 的条件. ■

13.5.2 一个有较高重数的根

在 11.7.3 节, 当 q 个较小的总体根相等时, 我们用 q 个较小的样本根的渐近分布. 我们现在导出这个分布. 令

$$\boldsymbol{\Lambda}=\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \lambda^*\boldsymbol{I}_q\end{pmatrix}, \tag{12}$$

其中对角矩阵 $\boldsymbol{\Lambda}_1$ 的对角元是不相等的, 并且大于 $\lambda^*(>0)$. 令

$$\boldsymbol{T}=\begin{pmatrix}\boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{pmatrix},\qquad \boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}_{11} & \boldsymbol{Y}_{12}\\ \boldsymbol{Y}_{21} & \boldsymbol{Y}_{22}\end{pmatrix},\qquad \boldsymbol{L}=\begin{pmatrix}\boldsymbol{L}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{L}_2\end{pmatrix}. \tag{13}$$

则 $\boldsymbol{T}\xrightarrow{p}\boldsymbol{\Lambda}$, 暗含了 $\boldsymbol{L}\xrightarrow{p}\boldsymbol{\Lambda}$, $\boldsymbol{Y}_{11}\xrightarrow{p}\boldsymbol{I}, \boldsymbol{Y}_{12}\xrightarrow{p}\mathbf{0}, \boldsymbol{Y}_{21}\xrightarrow{p}\mathbf{0}$, 但是 $\boldsymbol{Y}_{22}$ 并没有一个概率极限. 尽管如此还是有, $\boldsymbol{Y}_{22}\boldsymbol{Y}_{22}'\xrightarrow{p}\boldsymbol{I}_q$. 令 $\boldsymbol{Y}_{22}$ 的奇异值分解是 $\boldsymbol{EJF}$, 其中 $\boldsymbol{J}$ 是对角的, $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是正交的. 定义 $\boldsymbol{C}_2=\boldsymbol{EF}$, 是正交的. 令 $\boldsymbol{U}=\sqrt{n}(\boldsymbol{I}-\boldsymbol{\Lambda}), \boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda})$ 有类似 $\boldsymbol{T}$ 和 $\boldsymbol{L}$ 的划分. 定义 $\boldsymbol{W}_{11}=\sqrt{n}(\boldsymbol{Y}_{11}-\boldsymbol{I}), \boldsymbol{W}_{12}=\sqrt{n}\boldsymbol{Y}_{12}, \boldsymbol{W}_{21}=\sqrt{n}\boldsymbol{Y}_{21}, \boldsymbol{W}_{22}=\sqrt{n}(\boldsymbol{Y}_{22}=\boldsymbol{C}_2)=\sqrt{n}\boldsymbol{E}(\boldsymbol{J}-\boldsymbol{I}_q)\boldsymbol{F}$. 则 (4) 式可以写成

$$\begin{aligned}
&\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \lambda^*\boldsymbol{I}_q\end{pmatrix}+\frac{1}{\sqrt{n}}\begin{pmatrix}\boldsymbol{U}_{11} & \boldsymbol{U}_{12}\\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22}\end{pmatrix} \qquad (14)\\
=&\left[\begin{pmatrix}\boldsymbol{I}_{p-q} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}_2\end{pmatrix}+\frac{1}{\sqrt{n}}\begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{W}_{12}\\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22}\end{pmatrix}\right]\\
&\cdot\left[\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \lambda^*\boldsymbol{I}_q\end{pmatrix}+\frac{1}{\sqrt{n}}\begin{pmatrix}\boldsymbol{D}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{D}_2\end{pmatrix}\right]\\
&\cdot\left[\begin{pmatrix}\boldsymbol{I}_{p-q} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}_2'\end{pmatrix}+\frac{1}{\sqrt{n}}\begin{pmatrix}\boldsymbol{W}_{11}' & \boldsymbol{W}_{12}'\\ \boldsymbol{W}_{21}' & \boldsymbol{W}_{22}'\end{pmatrix}\right]\\
=&\begin{pmatrix}\boldsymbol{\Lambda}_1 & \mathbf{0}\\ \mathbf{0} & \lambda^*\boldsymbol{I}_q\end{pmatrix}+\frac{1}{\sqrt{n}}\left[\begin{pmatrix}\boldsymbol{D}_1 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{C}_2\boldsymbol{D}_2\boldsymbol{C}_2'\end{pmatrix}\right.\\
&\left.+\begin{pmatrix}\boldsymbol{W}_{11}\boldsymbol{\Lambda}_1 & \lambda^*\boldsymbol{W}_{12}\boldsymbol{C}_2'\\ \boldsymbol{W}_{21}\boldsymbol{\Lambda}_1 & \lambda^*\boldsymbol{W}_{22}\boldsymbol{C}_2'\end{pmatrix}+\begin{pmatrix}\boldsymbol{\Lambda}_1\boldsymbol{W}_{11}' & \boldsymbol{\Lambda}_1\boldsymbol{W}_{21}'\\ \lambda^*\boldsymbol{C}_2\boldsymbol{W}_{12}' & \lambda^*\boldsymbol{C}_2\boldsymbol{W}_{22}'\end{pmatrix}\right]+\frac{1}{n}\boldsymbol{M},
\end{aligned}$$

其中 $\boldsymbol{M}$ 的子矩阵是 $\boldsymbol{C}_2, \boldsymbol{\Lambda}_1, \lambda^*\boldsymbol{I}_q, \boldsymbol{D}_k, \boldsymbol{W}_{k1}, 1/\sqrt{n}$ 的乘积的和. $\boldsymbol{Y}(\boldsymbol{I}_p=\boldsymbol{YY}')$ 的正交性包括了

$$\boldsymbol{I}_p=\begin{pmatrix}\boldsymbol{I}_{p-q} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_q\end{pmatrix}+\frac{1}{\sqrt{n}}\left[\begin{pmatrix}\boldsymbol{W}_{11} & \boldsymbol{W}_{12}\boldsymbol{C}_2'\\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22}\boldsymbol{C}_2'\end{pmatrix}+\begin{pmatrix}\boldsymbol{W}_{11}' & \boldsymbol{W}_{21}'\\ \boldsymbol{C}_2'\boldsymbol{W}_{12}' & \boldsymbol{C}_2\boldsymbol{W}_{22}'\end{pmatrix}\right]+\frac{1}{n}\boldsymbol{N}, \tag{15}$$

其中 $\boldsymbol{N}$ 的子矩阵是 $\boldsymbol{W}_{kl}$ 的乘积的和. 从 (14) 和 (15) 我们得到

$$\boldsymbol{U}_{22}=\boldsymbol{C}_2\boldsymbol{D}_2\boldsymbol{C}_2'+O_p(1/\sqrt{n}). \tag{16}$$

$(1/\lambda^*)\boldsymbol{U}_{22}$ 的极限分布有 13.3 节 (25) 式 p 被 q 取代后的密度. $\boldsymbol{D}_2$ 和 $\boldsymbol{C}_2$ 的极限分布是由 $\boldsymbol{U}_{22}^* = \boldsymbol{Y}_{22}^*\boldsymbol{D}_2^*\boldsymbol{Y}_{22}^{*\prime}$ 定义的 $\boldsymbol{D}_2^*$ 和 $\boldsymbol{Y}_{22}^*$ 的极限分布, 其中 $(1/\lambda^*)\boldsymbol{U}_{22}^*$ 有 13.3 节的密度 (25) 式.

定理 13.5.2 在定理 13.5.1 的条件下, $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\Lambda}_1, \lambda^*\boldsymbol{I}_q)$. 则 $d_{p-q+1}, \cdots, d_p$ 的极限分布的密度是

$$2^{-\frac{1}{2}q}(\lambda^*\pi)^{q(q-1)/4}\Gamma_p^{-\frac{1}{2}}\left(\frac{1}{2}p\right)\exp\left(-\frac{1}{2\lambda^*}\sum_{i=p-q+1}^{p} d_i^2\right)\prod_{i<j}(d_i-d_j). \tag{17}$$

为了证明前面的推导, 我们注意到 $\boldsymbol{D}_2$ 和 $\boldsymbol{Y}_{22}$ 是 $\boldsymbol{U}$ 的函数, $\boldsymbol{U}$ 依靠于 n 并且收敛到 $\boldsymbol{U}_{22}^* = \boldsymbol{Y}_{22}^*\boldsymbol{D}_{22}^*\boldsymbol{Y}_{22}^{*\prime}$ 的解. 我们可以用下面由 Anderson(1963a) 给出, 归功于 Rubin 的定理.

定理 13.5.3 令 $F_n(\boldsymbol{u})$ 是一个随机矩阵 $\boldsymbol{U}_n$ 的累积分布函数. 令 $\boldsymbol{V}_n$ 是 $\boldsymbol{U}_n$ 和 $\boldsymbol{V}_n = \boldsymbol{f}_n(\boldsymbol{u}_n)$ 的矩阵值函数, 令 $\boldsymbol{G}_n(\boldsymbol{v})$ 是 (诱导的)$\boldsymbol{V}_n$ 的分布. 假设在每一个 $F(\boldsymbol{u})$ 的连续点 $F_n(\boldsymbol{u}) \to F(\boldsymbol{u})$, 假设对于 $f(\boldsymbol{u})$ 的每个连续点 $\boldsymbol{u}$, 当 $\boldsymbol{u}_n \to \boldsymbol{u}$ 时 $f_n(\boldsymbol{u}_n) \to f(\boldsymbol{u})$. 令 $\boldsymbol{G}(\boldsymbol{v})$ 是随机矩阵 $\boldsymbol{V} = f(\boldsymbol{U})$ 的分布, 其中 $\boldsymbol{U}$ 有分布 $F(\boldsymbol{u})$, 如果根据 $F(\boldsymbol{u})$, $f(\boldsymbol{u})$ 的不连续集的概率是 0, 则在 $\boldsymbol{G}(\boldsymbol{v})$ 的每个连续点,

$$\lim_{n\to\infty} \boldsymbol{G}_n(\boldsymbol{v}) = \boldsymbol{G}(\boldsymbol{v}). \tag{18}$$

证明 $\boldsymbol{U}(n)$ 和

$$(\boldsymbol{D}_2(n), \boldsymbol{Y}_{22}(n)) = f_n(U(nn)) \tag{19}$$

满足定理条件的细节, 由 Anderson(1963a) 给出.

13.6 有两个 Wishart 矩阵情况下的渐近分布

13.6.1 所有总体的根都不相同

13.2 节我们研究了

$$|\boldsymbol{S}^* - l\boldsymbol{T}^*| = 0 \tag{1}$$

的根 $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_p$ 的分布, 满足

$$(\boldsymbol{S}^* - l\boldsymbol{T}^*)\boldsymbol{x}^* = \boldsymbol{0} \tag{2}$$

的向量, 当 $\boldsymbol{A}^* = m\boldsymbol{S}^*$ 和 $\boldsymbol{B}^* = n\boldsymbol{T}^*$ 各自服从 $W(\boldsymbol{\Sigma}, m)$ 和 $W(\boldsymbol{\Sigma}, n)$ 分布且独立时, $\boldsymbol{x}^{*\prime}\boldsymbol{T}^*\boldsymbol{x}^* = 1$. 本节我们讨论当 $\boldsymbol{A}^*$ 和 $\boldsymbol{B}^*$ 各自服从 $W(\Phi, m)$ 和 $W(\boldsymbol{\Sigma}, n)$ 分布且独立时, $n \to \infty$ 和 $m/n \to \eta > 0$ 时根和向量的渐近分布. 我们假定

$$|\Phi - \lambda\boldsymbol{\Sigma}| = 0 \tag{3}$$

的根是不相同的. (13.2 节 $\lambda_1 = \cdots = \lambda_p = 1$.)

定理 13.6.1 令 $m\boldsymbol{S}^*$ 和 $n\boldsymbol{T}^*$ 是独立的, 各自服从 $W(\Phi, m)$ 和 $W(\boldsymbol{\Sigma}, n)$ 分布. 令 $\lambda_1 > \lambda_2 > \cdots > \lambda_p (> 0)$ 是 (3) 的根, 令 $\boldsymbol{\Lambda}$ 是对角矩阵, 其对角元是按递减次序排列的根; 令 $\gamma_1, \cdots, \gamma_p$ 是

$$(\boldsymbol{\Phi} - \lambda_i \boldsymbol{\Sigma})\boldsymbol{\gamma} = \mathbf{0}, \quad i = 1, \cdots, p \tag{4}$$

的解. $\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma} = 1$, $\gamma_{1i} \geqslant 0$, 令 $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_p)$. 令 $l_1 \geqslant \cdots \geqslant l_p (> 0)$ 是 (1) 的根, 令 $\boldsymbol{L}$ 是对角矩阵, 其对角元是按递减次序排列的根; 令 $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_p^*$ 是 $l = l_i, i = 1, \cdots, p, \boldsymbol{x}^{*'}\boldsymbol{T}^*\boldsymbol{x}^* = 1$ 和 $x_{1i}^* > 0$ 时 (2) 式的根, 令 $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_p^*)$. 定义 $\boldsymbol{Z}^* = \sqrt{n}(\boldsymbol{X}^* - \boldsymbol{\Gamma})$ 和对角矩阵 $\boldsymbol{D} = \sqrt{n}(\boldsymbol{L} - \boldsymbol{\Lambda})$. 则 $\boldsymbol{D}$ 和 $\boldsymbol{Z}^*$ 的极限分布是正态的, 在 $n \to \infty, m \to \infty, m/n \to \eta (> 0)$ 时有均值是 0. 渐近方差和协方差不为零

$$\mathrm{AVar}(d_i) = 2\frac{\lambda_i^2(1+\eta)}{\eta}, \tag{5}$$

$$\mathrm{AVar}(\boldsymbol{z}_i^*) = \sum_{\substack{k=1\\k\neq i}}^{p} \frac{\lambda_i(\lambda_k + \eta\lambda_i)}{\eta(\lambda_k - \lambda_i)^2}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k' + \frac{1}{2}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i', \tag{6}$$

$$\mathrm{AVar}(\boldsymbol{d}_i, \boldsymbol{z}_i^*) = \lambda_i\boldsymbol{\gamma}_i, \tag{7}$$

$$\mathrm{AVar}(\boldsymbol{z}_i^*, \boldsymbol{z}_j^*) = -\frac{\lambda_i\lambda_j(1+\eta)}{\eta(\lambda_j - \lambda_i)^2}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_i', \quad i \neq j. \tag{8}$$

证明　令

$$\boldsymbol{S} = \boldsymbol{\Gamma}'\boldsymbol{S}^*\boldsymbol{\Gamma}, \quad \boldsymbol{T} = \boldsymbol{\Gamma}'\boldsymbol{T}^*\boldsymbol{\Gamma}. \tag{9}$$

则 $m\boldsymbol{S}$ 和 $n\boldsymbol{T}$ 是独立地服从 $W(\boldsymbol{\Lambda}, m)$ 和 $W(\boldsymbol{I}, n)$ 分布的 (7.3.3 节). 那么 $l_1, \cdots, l_p$ 是

$$|\boldsymbol{S} - l\boldsymbol{T}| = 0 \tag{10}$$

的根. 令 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p$ 是

$$(\boldsymbol{S} - l_i\boldsymbol{T})\boldsymbol{x} = \mathbf{0}, \quad i = 1, \cdots, p \tag{11}$$

的解. $\boldsymbol{x}'\boldsymbol{T}\boldsymbol{x} = 1$, 令 $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p)$. 则 $\boldsymbol{x}_i^* = \boldsymbol{\Gamma}\boldsymbol{x}_i$ 和 $\boldsymbol{X}^* = \boldsymbol{\Gamma}\boldsymbol{X}$ 除了 $\boldsymbol{X}$(或者 $\boldsymbol{X}^*$) 的列乘 -1 的可能性外. 如果 $\boldsymbol{Z} = \sqrt{n}(\boldsymbol{X} - \boldsymbol{I})$, 则 $\boldsymbol{Z}^* = \boldsymbol{\Gamma}\boldsymbol{Z}$(除了列乘 -1 的可能外).

我们现在寻找 $\boldsymbol{D}$ 和 $\boldsymbol{Z}$ 的极限分布. 令 $\sqrt{n}(\boldsymbol{S} - \boldsymbol{\Lambda}) = \boldsymbol{U}$ 和 $\sqrt{n}(\boldsymbol{T} - \boldsymbol{I}) = \boldsymbol{V}$. 则 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 有独立的极限正态分布, 其均值是 $\mathbf{0}$. $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 中函数独立的元在极限分布中是统计独立的. 方差是 $\mathrm{E}(u_{ii}^2) = 2(n/m)\lambda_i^2 \to 2\lambda_i^2/\eta$; $\mathrm{E}(u_{ij}^2) = (n/m)\lambda_i\lambda_j \to \lambda_i\lambda_j/\eta, i \neq j$; $\mathrm{E}(v_{ii}^2) = 2$; $\mathrm{E}(v_{ij}^2) = 1, i \neq j$.

从 $\boldsymbol{L}$ 和 $\boldsymbol{X}$ 的定义我们有 $\boldsymbol{S}\boldsymbol{X} = \boldsymbol{T}\boldsymbol{X}\boldsymbol{L}, \boldsymbol{X}'\boldsymbol{T}\boldsymbol{X} = \boldsymbol{I}, \boldsymbol{X}'\boldsymbol{S}\boldsymbol{X} = \boldsymbol{L}$. 如果我们令 $\boldsymbol{X}^{-1} = \boldsymbol{G}$, 我们得到了

$$\boldsymbol{S} = \boldsymbol{G}'\boldsymbol{L}\boldsymbol{G}, \qquad \boldsymbol{T} = \boldsymbol{G}'\boldsymbol{G}. \tag{12}$$

我们需要 $g_{ii} > 0, i = 1, \cdots, p$. 因为 $\boldsymbol{S} \xrightarrow{p} \boldsymbol{\Lambda}, \boldsymbol{T} \xrightarrow{p} \boldsymbol{I}$, 我们有 $\boldsymbol{L} \xrightarrow{p} \boldsymbol{\Lambda}$ 和 $\boldsymbol{G} \xrightarrow{p} \boldsymbol{I}$. 令 $\sqrt{n}(\boldsymbol{G} - \boldsymbol{I}) = \boldsymbol{H}$. 那么可将 (12) 式写成

$$\boldsymbol{\Lambda} = \frac{1}{\sqrt{n}}\boldsymbol{U} = \left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{H}'\right)\left(\boldsymbol{\Lambda} + \frac{1}{\sqrt{n}}\boldsymbol{D}\right)\left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{H}\right), \tag{13}$$

$$\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{V} = \left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{H}'\right)\left(\boldsymbol{I} + \frac{1}{\sqrt{n}}\boldsymbol{H}\right). \tag{14}$$

这些可以写成

$$U = D + \Lambda H + H'\Lambda + \frac{1}{\sqrt{n}}(DH + H'D + H'\Lambda H) + \frac{1}{n}H'DH, \tag{15}$$

$$V = H + H' + \frac{1}{\sqrt{n}}H'H. \tag{16}$$

如果忽略了 $\frac{1}{\sqrt{n}}$ 阶和 $\frac{1}{n}$ 阶的项 (如 13.5 节), 我们可以写成

$$U = D + \Lambda H + H'\Lambda, \tag{17}$$

$$V = H + H', \tag{18}$$

$$U - V\Lambda = D + \Lambda H - H\Lambda. \tag{19}$$

(18) 的对角元和 (19) 的元是

$$v_{ii} = 2h_{ii}, \tag{20}$$

$$u_{ii} - \lambda_i v_{ii} = d_i, \tag{21}$$

$$u_{ij} - v_{ij}\lambda_j = (\lambda_i - \lambda_j)h_{ij}, \quad i \neq j. \tag{22}$$

$\boldsymbol{H}$ 和 $\boldsymbol{D}$ 的极限分布是正态的, 均值是 0. $\boldsymbol{H}$ 的非对角元素的对 (h_{ij}, h_{ji}) 是独立的, 有方差

$$\text{AVar}(h_{ij}) = \frac{\lambda_j(\lambda_i + \eta\lambda_j)}{\eta(\lambda_i - \lambda_j)^2}, \quad i \neq j, \tag{23}$$

协方差

$$\text{ACov}(h_{ij}, h_{ji}) = -\frac{\lambda_i\lambda_j(1+\eta)}{\eta(\lambda_i - \lambda_j)^2}, \quad i \neq j. \tag{24}$$

$\boldsymbol{D}$ 和 $\boldsymbol{H}$ 的对角元的对 (d_i, h_{ii}) 同 (5) 的方差是独立的,

$$\text{AVar}(h_{ii}) = \frac{1}{2}, \tag{25}$$

协方差是

$$\text{ACov}(d_i, h_{ii}) = -\lambda_i. \tag{26}$$

$\boldsymbol{D}$ 和 $\boldsymbol{H}$ 的对角元和 $\boldsymbol{H}$ 的非对角元是独立的.

$\boldsymbol{D}$ 和 $\boldsymbol{H}$ 的极限分布是正态的, 定理 4.2.3 给出了证明. $\boldsymbol{S}$ 和 $\boldsymbol{T}$ 是 $\boldsymbol{L}$ 和 $\boldsymbol{G}$ 的多项式, 它们的导数也是多项式, 因此是连续的. 因为方程 (12) 加上辅助的条件可以唯一地解出 $\boldsymbol{L}$ 和 $\boldsymbol{G}$, 逆函数在 $\boldsymbol{L} = \boldsymbol{\Lambda}$ 和 $\boldsymbol{G} = \boldsymbol{I}$ 同样也是连续可微的. 由定理 4.2.3, $\boldsymbol{D} = \sqrt{n}(\boldsymbol{L} - \boldsymbol{\Lambda})$ 和 $\boldsymbol{H} = \sqrt{n}(\boldsymbol{G} - \boldsymbol{I})$ 有极限正态分布. 进而, $\boldsymbol{X} = \boldsymbol{G}^{-1}$ 是在 $\boldsymbol{G} = \boldsymbol{I}, \boldsymbol{Z} = \sqrt{n}(\boldsymbol{X} - \boldsymbol{I}) = \sqrt{n}(\boldsymbol{G}^{-1} - \boldsymbol{I})$ 是连续可微的, 有极限分布 $-\boldsymbol{H}$. (展开 $\sqrt{n}\{[\boldsymbol{I} + (1/\sqrt{n})\boldsymbol{H}]^{-1} - \boldsymbol{I}\}$.) 因为以接近 1 的概率有 $\boldsymbol{G} \xrightarrow{p} \boldsymbol{I}, \boldsymbol{X} \xrightarrow{p} \boldsymbol{I}, x_{ii} > 0, i = 1, \cdots, p$. 则 $\boldsymbol{Z}^* = \sqrt{n}(\boldsymbol{X}^* - \boldsymbol{\Gamma})$ 有极限分布 $\boldsymbol{\Gamma Z}$. (因为 $\boldsymbol{X} \xrightarrow{p} \boldsymbol{I}$, 以接近 1 的概率有 $\boldsymbol{X}^* = \boldsymbol{\Gamma X} \xrightarrow{p} \boldsymbol{\Gamma}$ 和 $x_{1i} > 0, i = 1, \cdots, p$.) 渐近方差和协方差 (6) 和 (8) 式由 (23)~(26) 获得.

Anderson(1986b) 在另一个矩阵有任意多重总体根的情况下, 得到了样本协方差阵的特征根和特征向量的极限分布.

13.6.2 一个有较高重数的根

13.6.1 节假设 $m\boldsymbol{S}^*$ 和 $n\boldsymbol{T}^*$ 各自服从 $W(\boldsymbol{\Phi},m)$ 和 $W(\boldsymbol{\Sigma},n)$ 分布且独立, $|\boldsymbol{\Phi}-\lambda\boldsymbol{\Sigma}|=0$ 的根是不同的. 本节我们假设前 k 个较大的根是不相等的, $p-k$ 个较小的根假定相等. 令特征根的对角矩阵 $\boldsymbol{\Lambda}$ 是 $\boldsymbol{\Lambda}=\mathrm{diag}(\boldsymbol{\Lambda}_1,\lambda^*\boldsymbol{I}_{p-k})$, 令 $\boldsymbol{\Gamma}$ 是满足

$$\boldsymbol{\Phi}\boldsymbol{\Gamma}=\boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\Lambda},\quad \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma}=\boldsymbol{I} \tag{27}$$

的矩阵. 用 (9) 式定义 $\boldsymbol{S}$ 和 $\boldsymbol{T}$, (12) 式定义对角矩阵 $\boldsymbol{L}$ 和 $\boldsymbol{G}$. 则 $\boldsymbol{S}\xrightarrow{p}\boldsymbol{\Lambda},\boldsymbol{T}\xrightarrow{p}\boldsymbol{I}_p,\boldsymbol{L}\xrightarrow{p}\boldsymbol{\Lambda}$. 将 $\boldsymbol{S},\boldsymbol{T},\boldsymbol{L}$ 和 $\boldsymbol{G}$ 分成

$$\boldsymbol{S}=\begin{bmatrix}\boldsymbol{S}_{11} & \boldsymbol{S}_{12}\\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22}\end{bmatrix},\quad \boldsymbol{T}=\begin{bmatrix}\boldsymbol{T}_{11} & \boldsymbol{T}_{12}\\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22}\end{bmatrix},$$
$$\boldsymbol{L}=\begin{bmatrix}\boldsymbol{L}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}_2\end{bmatrix},\quad \boldsymbol{G}=\begin{bmatrix}\boldsymbol{G}_{11} & \boldsymbol{G}_{12}\\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22}\end{bmatrix}, \tag{28}$$

其中 $\boldsymbol{S}_{11},\boldsymbol{T}_{11},\boldsymbol{L}_1$ 和 $\boldsymbol{G}_{11}$ 是 $k\times k$ 的. 则 $\boldsymbol{G}_{11}\xrightarrow{p}\boldsymbol{I}_k,\boldsymbol{G}_{12}\xrightarrow{p}\boldsymbol{0},\boldsymbol{G}_{21}\xrightarrow{p}\boldsymbol{0}$, 但是 $\boldsymbol{G}_{22}$ 没有概率极限. 代替有 $\boldsymbol{G}_{22}'\boldsymbol{G}_{22}\xrightarrow{p}\boldsymbol{I}_{p-k}$. 令 $\boldsymbol{G}_{22}$ 的奇异值分解是 $\boldsymbol{EJF}$, 其中 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是正交的, $\boldsymbol{J}$ 是对角的. 令 $\boldsymbol{C}_2=\boldsymbol{EF}$.

$\boldsymbol{U}=\sqrt{n}(\boldsymbol{S}-\boldsymbol{\Lambda})$ 和 $\boldsymbol{V}=\sqrt{n}(\boldsymbol{T}-\boldsymbol{I})$ 的极限分布是正态的, 其协方差结构由 (12) 式给出, $\lambda_{k+1}=\cdots=\lambda_p=\lambda^*$, 定义 $\boldsymbol{D}=\sqrt{n}(\boldsymbol{L}-\boldsymbol{\Lambda}),\boldsymbol{H}_{11}=\sqrt{n}(\boldsymbol{G}_{11}-\boldsymbol{I}),\boldsymbol{H}_{12}=\sqrt{n}\boldsymbol{G}_{12},\boldsymbol{H}_{21}=\sqrt{n}\boldsymbol{G}_{21},\boldsymbol{H}_{22}=\sqrt{n}(\boldsymbol{G}_{22}-\boldsymbol{C}_2)=\sqrt{n}\boldsymbol{E}(\boldsymbol{J}-\boldsymbol{I}_{p-k})\boldsymbol{F}$. 则 (13) 和 (15) 可以被

$$\begin{aligned}
&\begin{bmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \lambda^*\boldsymbol{I}_{p-k}\end{bmatrix}+\frac{1}{\sqrt{n}}\begin{bmatrix}\boldsymbol{U}_{11} & \boldsymbol{U}_{12}\\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22}\end{bmatrix} \qquad (29)\\
=&\begin{bmatrix}\boldsymbol{I}_k+\frac{1}{\sqrt{n}}\boldsymbol{H}_{11}' & \frac{1}{\sqrt{n}}\boldsymbol{H}_{21}'\\ \frac{1}{\sqrt{n}}\boldsymbol{H}_{12}' & \boldsymbol{C}_2'+\frac{1}{\sqrt{n}}\boldsymbol{H}_{22}'\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_1+\frac{1}{\sqrt{n}}\boldsymbol{D}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \lambda^*\boldsymbol{I}_{p-k}+\frac{1}{\sqrt{n}}\boldsymbol{D}_2\end{bmatrix}\\
&\cdot\begin{bmatrix}\boldsymbol{I}_k+\frac{1}{\sqrt{n}}\boldsymbol{H}_{11} & \frac{1}{\sqrt{n}}\boldsymbol{H}_{12}\\ \frac{1}{\sqrt{n}}\boldsymbol{H}_{21} & \boldsymbol{C}_2+\frac{1}{\sqrt{n}}\boldsymbol{H}_{22}\end{bmatrix}\\
=&\begin{bmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \lambda^*\boldsymbol{I}_{p-k}\end{bmatrix}+\frac{1}{\sqrt{n}}\begin{bmatrix}\boldsymbol{D}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_2'\boldsymbol{D}_2\boldsymbol{C}_2\end{bmatrix}+\frac{1}{\sqrt{n}}\begin{bmatrix}\boldsymbol{\Lambda}_1\boldsymbol{H}_{11} & \boldsymbol{\Lambda}_1\boldsymbol{H}_{12}\\ \lambda^*\boldsymbol{C}_2'\boldsymbol{H}_{21} & \lambda^*\boldsymbol{C}_2'\boldsymbol{H}_{22}\end{bmatrix}\\
&+\frac{1}{\sqrt{n}}\begin{bmatrix}\boldsymbol{H}_{11}'\boldsymbol{\Lambda}_1 & \lambda^*\boldsymbol{H}_{21}'\boldsymbol{C}_2\\ \boldsymbol{H}_{12}'\boldsymbol{\Lambda}_1 & \lambda^*\boldsymbol{H}_{22}'\boldsymbol{C}_2\end{bmatrix}+O_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

取代, (14) 和 (16) 可以被

$$\begin{bmatrix} I_k & 0 \\ 0 & I_{p-k} \end{bmatrix} + \frac{1}{\sqrt{n}} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{30}$$
$$= \begin{bmatrix} I + \frac{1}{\sqrt{n}} H'_{11} & \frac{1}{\sqrt{n}} H'_{21} \\ \frac{1}{\sqrt{n}} H'_{12} & C'_2 + \frac{1}{\sqrt{n}} H'_{22} \end{bmatrix} \begin{bmatrix} I + \frac{1}{\sqrt{n}} H_{11} & \frac{1}{\sqrt{n}} H_{12} \\ \frac{1}{\sqrt{n}} H_{21} & C_2 + \frac{1}{\sqrt{n}} H_{22} \end{bmatrix}$$
$$= \begin{bmatrix} I & 0 \\ 0 & I_{p-k} \end{bmatrix} + \frac{1}{\sqrt{n}} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & C'_2 H_{22} \end{bmatrix} + \frac{1}{\sqrt{n}} \begin{bmatrix} H'_{11} & H'_{21} \\ H'_{12} & H'_{22} C_2 \end{bmatrix} + \frac{1}{n} H'H$$

取代. 如果我们忽略了阶 $1/\sqrt{n}$ 和 $1/n$ 的项, 代替 (19) 式我们可以写成

$$\begin{bmatrix} U_{11} - V_{11}\Lambda_1 & U_{12} - \lambda^* V_{12} \\ U_{21} - V_{21}\Lambda_1 & U_{22} - \lambda^* V_{22} \end{bmatrix} = \begin{bmatrix} D_1 + \Lambda_1 H_{11} - H_{11}\Lambda_1 & (\lambda^* I - \Lambda_1) H_{12} C_2 \\ C'_2 H_{21} (\lambda^* I - V_1) & C'_2 D_2 C_2 \end{bmatrix}. \tag{31}$$

则 $v_{ii} = 2h_{ii}, i = 1, \cdots, k; u_{ii} - \lambda_i v_{ii} = d_i, i = 1, \cdots, k; u_{ij} - v_{ij}\lambda_j = (\lambda_i - \lambda_j) h_{ij}, i \neq j, i, j = 1, \cdots, k$; $U_{22} - \lambda^* V_{22} = C'_2 D_2 C_2; C_2(U_{21} - V_{21}\Lambda_1) = H_{21}(\lambda^* I - \Lambda_1)$; $(U_{12} - \lambda^* V_{12}) C'_2 = (\lambda^* I - \Lambda_1) H_{12}$. $U_{22} - \lambda^* V_{22}$ 的极限分布是正态的, 有均值 $\mathbf{0}$; $\mathrm{E}(u_{ii} - \lambda^* v_{ii})^2 = \mathrm{E}(d_i^2) = 2\lambda^{*2}(1+\eta)/\eta, i = k+1, \cdots, p$; $\mathrm{E}(u_{ij} - \lambda^* v_{ij})^2 = \lambda^{*2}(1+\eta)/\eta, i \neq j, i, j = k+1, \cdots, p$. D_2 和 C_2 的极限分布是由 $U_{22} - \lambda^* V_{22} = C'_2 D_2 C_2$ 定义的 D_2 和 C_2 的分布, 其中 $(1/\lambda^*)(U_{22} - V_{22})$ 有 13.3 节的密度 (25).

13.7 一个回归模型下的渐近分布

13.7.1 两组变量均随机

样本典型相关 $l_1, \cdots, l_{p_2}$, 向量 $\hat{\alpha}_1, \cdots, \hat{\alpha}_{p_1}$ 和 $\hat{\gamma}_1, \cdots, \hat{\gamma}_{p_2}$ 由 12.3 节定义. 集合 $\hat{\gamma}_1, \cdots, \hat{\gamma}_{p_2}$ 和 $l_1, \cdots, l_{p_2}$ 由

$$S_{21} S_{11}^{-1} S_{12} \hat{\gamma} = S_{22} \hat{\gamma} l^2, \quad \hat{\gamma}' S_{22} \hat{\gamma} = 1 \tag{1}$$

定义. 当 $X = (X^{(1)\prime}, X^{(2)\prime})'$ 有正态分布时, 以及当 $X^{(1)}$ 有正态分布, 且其期望值是非随机向量 $X^{(2)}$ 的线性函数时, 这些量的渐近分布由 Anderson(1999a) 给出. 我们现在寻找当 X 有正态分布时的渐近分布. 模型的回归形式是

$$X^{(1)} = \boldsymbol{\beta} X^{(2)} + Z, \tag{2}$$

其中 $X^{(2)}$ 和 Z 是独立的和正态分布的, 有期望值 $\mathrm{E}(X^{(2)}) = \mathbf{0}$ 和 $\mathrm{E}(Z) = \mathbf{0}$, 和协方差 $\mathrm{E}(X^{(2)} X^{(2)\prime}) = \Sigma_{22}, \mathrm{E}(ZZ') = \Sigma_{ZZ} (\mathrm{E}(X^{(2)} Z') = \mathbf{0})$. 则 $\mathrm{E}(X^{(1)}) = \mathbf{0}$, $\mathrm{E}(X^{(1)} X^{(1)\prime}) = \Sigma_{11} = \Sigma_{ZZ} + \boldsymbol{\beta}\Sigma_{22}\boldsymbol{\beta}', \mathrm{E}(X^{(1)} X^{(2)\prime}) = \boldsymbol{\beta}\Sigma_{22}$. 推断基于样本 X 的 n 个观测.

首先我们变换到典型变量 $U = A' X^{(1)}, V = \Gamma' X^{(2)}, W = A' Z$. 则 (1) 式变换到

$$\boldsymbol{U}=\boldsymbol{\Theta V}+\boldsymbol{W}, \tag{3}$$

其中 $\boldsymbol{\Theta}=\boldsymbol{A}'\boldsymbol{\beta}(\boldsymbol{\Gamma}')^{-1}, \mathrm{E}(\boldsymbol{UU}')=\boldsymbol{\Sigma}_{UU}=\boldsymbol{I}_{p_1}, \mathrm{E}(\boldsymbol{VV}')=\boldsymbol{\Sigma}_{VV}=\boldsymbol{I}_{p_2}, \mathrm{E}(\boldsymbol{UV}')=\boldsymbol{\Sigma}_{UV}=(\boldsymbol{\Lambda},\ 0)=\bar{\boldsymbol{\Lambda}}, \mathrm{E}(\boldsymbol{WW}')=\boldsymbol{\Sigma}_{WW}=\boldsymbol{I}_{p_1}-\boldsymbol{\Lambda}^2, \mathrm{E}(\boldsymbol{VW}')=\boldsymbol{0}$. [见 12.2 节的 (33)~(37) 和 (45).] 令样本协方差阵是 $\boldsymbol{S}_{UU}=\boldsymbol{A}'\boldsymbol{S}_{11}\boldsymbol{A}', \boldsymbol{S}_{UV}=\boldsymbol{A}'\boldsymbol{S}_{12}\boldsymbol{\Gamma}, \boldsymbol{S}_{VV}=\boldsymbol{\Gamma}'\boldsymbol{S}_{22}\boldsymbol{\Gamma}$. 令样本向量包括 $\boldsymbol{H}=\boldsymbol{\Gamma}^{-1}\hat{\boldsymbol{\Gamma}}=\boldsymbol{\Gamma}^{-1}(\hat{\gamma}_1,\cdots,\hat{\gamma}_{p_2})$. 则 $\boldsymbol{H}$ 满足

$$\boldsymbol{S}_{VU}\boldsymbol{S}_{UU}^{-1}\boldsymbol{S}_{UV}\boldsymbol{H}=\boldsymbol{S}_{VV}\boldsymbol{H}\hat{\boldsymbol{\Lambda}}^{+2}, \qquad \boldsymbol{H}'\boldsymbol{S}_{VV}\boldsymbol{H}=\boldsymbol{I}_{p_2}, \tag{4}$$

其中 $\hat{\boldsymbol{\Lambda}}^{+}=\mathrm{diag}(\hat{\lambda}_1,\cdots,\hat{\lambda}_{p_1},0,\cdots,0)$; 如果 $p_1<p_2$, 则 $\hat{\boldsymbol{\Lambda}}^{+}$ 中有 p_2-p_1 个 0.

我们有 $\boldsymbol{S}_{UU}\xrightarrow{p}\boldsymbol{I}_{p_1}, \boldsymbol{S}_{VV}\xrightarrow{p}\boldsymbol{I}_{p_2}, \boldsymbol{S}_{UV}\xrightarrow{p}\bar{\boldsymbol{\Lambda}}$. 则 $\hat{\boldsymbol{\Lambda}}=\mathrm{diag}(\hat{\lambda}_1,\cdots,\hat{\lambda}_{p_1})\xrightarrow{p}\boldsymbol{\Lambda}$. 令

$$\boldsymbol{H}=(\boldsymbol{H}_1,\ \boldsymbol{H}_2)=\begin{bmatrix}\boldsymbol{H}_{11} & \boldsymbol{H}_{12}\\ \boldsymbol{H}_{21} & \boldsymbol{H}_{22}\end{bmatrix}, \tag{5}$$

其中 $\boldsymbol{H}_{11}$ 是 $p_1\times p_1$ 的, $\boldsymbol{H}_{22}$ 是 $(p_2-p_1)\times(p_2-p_1)$. (4) 式的前 p_1 列是

$$\boldsymbol{S}_{VU}\boldsymbol{S}_{UU}^{-1}\boldsymbol{S}_{UV}\boldsymbol{H}_1=\boldsymbol{S}_{VV}\boldsymbol{H}_1\hat{\boldsymbol{\Lambda}}^2; \tag{6}$$

(4) 式的后 p_2-p_1 列是 $\boldsymbol{S}_{UV}\boldsymbol{H}_2=\boldsymbol{0}$. 则 $\boldsymbol{H}_{11}\xrightarrow{p}\boldsymbol{I}_{p_1}, \boldsymbol{H}_{12}\xrightarrow{p}\boldsymbol{0}, \boldsymbol{H}_{21}\xrightarrow{p}\boldsymbol{0}$. 但是 (4) 式的概率极限仅仅包含了 $\boldsymbol{H}_{22}'\boldsymbol{H}_{22}\xrightarrow{p}\boldsymbol{I}_{p_2-p_1}$. 令 $\boldsymbol{H}_{22}$ 的奇异值分解是 $\boldsymbol{H}_{22}=\boldsymbol{EJF}$.

定义 $\boldsymbol{S}_{UU}^*=\sqrt{n}(\boldsymbol{S}_{UU}-\boldsymbol{I}_{p_1}), \boldsymbol{S}_{VV}^*=\sqrt{n}(\boldsymbol{S}_{VV}-\boldsymbol{I}_{p_2}), \boldsymbol{S}_{UV}^*=\sqrt{n}(\boldsymbol{S}_{UV}-\bar{\boldsymbol{\Lambda}}), \boldsymbol{H}_1^*=\sqrt{n}(\boldsymbol{H}_1-\boldsymbol{I}_{(p_1)}), \bar{\boldsymbol{\Lambda}}^*=[\sqrt{n}(\hat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda}),\boldsymbol{0}]$, 其中 $\boldsymbol{I}_{(p_1)}=(\boldsymbol{I}_{p_1},\boldsymbol{0})'$. 则把 (6) 展开后可得

$$\begin{aligned}&\left(\bar{\boldsymbol{\Lambda}}'+\frac{1}{\sqrt{n}}\boldsymbol{S}_{VU}^*\right)\left(\boldsymbol{I}_{p_1}+\frac{1}{\sqrt{n}}\boldsymbol{S}_{UU}^*\right)^{-1}\left(\bar{\boldsymbol{\Lambda}}+\frac{1}{\sqrt{n}}\boldsymbol{S}_{UV}^*\right)\left(\boldsymbol{I}_{(p_1)}+\frac{1}{\sqrt{n}}\boldsymbol{H}_1^*\right)\\ &=\left(\boldsymbol{I}_{p_2}+\frac{1}{\sqrt{n}}\boldsymbol{S}_{VV}^*\right)\left(\boldsymbol{I}_{(p_1)}+\frac{1}{\sqrt{n}}\boldsymbol{H}_1^*\right)\left(\boldsymbol{\Lambda}+\frac{1}{\sqrt{n}}\boldsymbol{\Lambda}^*\right)^2.\end{aligned} \tag{7}$$

从 (7) 式我们得到

$$\begin{aligned}&\boldsymbol{S}_{VU}^*\bar{\boldsymbol{\Lambda}}\boldsymbol{I}_{(p_1)}-\bar{\boldsymbol{\Lambda}}'\boldsymbol{S}_{UU}^*\bar{\boldsymbol{\Lambda}}\boldsymbol{I}_{(p_1)}+\bar{\boldsymbol{\Lambda}}'\boldsymbol{S}_{UV}^*\boldsymbol{I}_{(p_1)}+\bar{\boldsymbol{\Lambda}}'\bar{\boldsymbol{\Lambda}}\boldsymbol{H}_1^*\\ &=\boldsymbol{S}_{VV}^*\boldsymbol{I}_{(p_1)}\boldsymbol{\Lambda}^2+\boldsymbol{H}_1^*\boldsymbol{\Lambda}^2+2\boldsymbol{I}_{(p_1)}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^*+o_p(1).\end{aligned} \tag{8}$$

从 $\boldsymbol{H}_1'\boldsymbol{S}_{VV}\boldsymbol{H}_1=\boldsymbol{I}_{p_1}$ 我们得到

$$\boldsymbol{H}_{11}^*+\boldsymbol{H}_{11}^{*'}=-\boldsymbol{S}_{VV}^{*11}+o_p(1). \tag{9}$$

用划分矩阵 (8) 的方法得到

$$\begin{bmatrix}\boldsymbol{S}_{UV}^{*11}\boldsymbol{\Lambda}-\boldsymbol{\Lambda}\boldsymbol{S}_{UU}^{*11}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{S}_{UV}^{*11}-\boldsymbol{S}_{VV}^{*11}\boldsymbol{\Lambda}^2\\ \boldsymbol{S}_{VU}^{*21}\boldsymbol{\Lambda}-\boldsymbol{S}_{VV}^{*21}\boldsymbol{\Lambda}^2\end{bmatrix}=\begin{bmatrix}2\boldsymbol{\Lambda}\boldsymbol{\Lambda}^*+\boldsymbol{H}_{11}^*\boldsymbol{\Lambda}^2-\boldsymbol{\Lambda}^2\boldsymbol{H}_{11}^*\\ \boldsymbol{H}_{21}^*\boldsymbol{\Lambda}^2\end{bmatrix}+o_p(1). \tag{10}$$

(10) 式的低阶子矩阵方程 $[(p_2-p_1)\times p_1]$ 是

$$\boldsymbol{H}_{21}^*\boldsymbol{\Lambda}=\boldsymbol{S}_{VU}^{*21}-\boldsymbol{S}_{VV}^{*21}\boldsymbol{\Lambda}+o_p(1)=\boldsymbol{S}_{VW}^{*21}+o_p(1)=(\boldsymbol{S}_{WV}^{*12})'+o_p(1). \tag{11}$$

(10) 式的上三角子矩阵方程的对角元是

$$\lambda_i^*=\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}[(u_{i\alpha}v_{i\alpha}-\lambda_i)-\frac{1}{2}\lambda_i(u_{ij}^2-1)-\frac{1}{2}\lambda_i(v_{i\alpha}^2-1)]+o_p(1). \tag{12}$$

(12) 式的右端是 $u_{i\alpha}$ 和 $v_{i\alpha}$ 样本相关系数的展开. 见 4.2.3 节. λ_i^* 的极限分布是 $N[0,(1-\lambda_i^2)^2]$.

(10) 中 $\boldsymbol{H}_{11}^*$ 的第 (i,j) 个元是

$$\begin{aligned}(\lambda_j^2-\lambda_i^2)h_{ij}^*=&\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}(\lambda_j v_{i\alpha}u_{j\alpha}+\lambda_i u_{i\alpha}v_{j\alpha}-\lambda_i\lambda_j u_{i\alpha}u_{j\alpha}\\&-\lambda_j^2 v_{i\alpha}v_{j\alpha})+o_p(1),\quad i\neq j,i,j=1,\cdots,p_1\end{aligned} \tag{13}$$

$(\lambda_j^2-\lambda_i^2)h_{ij}^*$ 和 $(\lambda_i^2-\lambda_j^2)h_{ji}^*$ 的渐近协方差是

$$\begin{bmatrix}(1-\lambda_j^2)(\lambda_i^2+\lambda_j^2-2\lambda_i^2\lambda_j^2) & (1-\lambda_i^2)(1-\lambda_j^2)(\lambda_i^2+\lambda_j^2)\\(1-\lambda_i^2)(1-\lambda_j^2)(\lambda_i^2+\lambda_j^2) & (1-\lambda_i^2)(\lambda_i^2+\lambda_j^2-2\lambda_i^2\lambda_j^2)\end{bmatrix}. \tag{14}$$

(h_{ij}^*,h_{ji}^2) 是和其他对不相关的.

假设 $p_1=p_2$. 则 $\hat{\boldsymbol{\Gamma}}^*=\boldsymbol{\Gamma}\boldsymbol{H}_{11}^*=\boldsymbol{\Gamma}\boldsymbol{H}^*$. 令 $\boldsymbol{\Gamma}=(\gamma_1,\cdots,\gamma_{p_1}),\hat{\boldsymbol{\Gamma}}=(\hat{\gamma}_1,\cdots,\hat{\gamma}_{p_1})$. 则 $\hat{\gamma}_j^*=\sum_{i=1}^{p_1}\hat{\gamma}_i\boldsymbol{h}_{ij}^*$, 其中的 h_{ij}^*, $i\neq j$, 由 (13) 式得到, h_{ii}^* 由 (9) 得到. 我们得到

$$n\mathrm{E}(\hat{\gamma}_j-\gamma_j)(\hat{\gamma}_j-\gamma_j)'=\frac{1}{2}\gamma_j\gamma_j'+(1-\lambda_j^2)\sum_{\substack{k=1\\k\neq j}}^{p_1}\frac{\lambda_k^2+\lambda_j^2-2\lambda_k^2\lambda_j^2}{(\lambda_j^2-\lambda_k^2)^2}, \tag{15}$$

$$n\mathrm{E}(\hat{\gamma}_j-\gamma_j)(\hat{\gamma}_l-\gamma_l)'=\frac{(1-\lambda_j^2)(1-\lambda_l^2)(\lambda_j^2+\lambda_l^2)}{(\lambda_j^2-\lambda_l^2)^2}\gamma_l\gamma_j',\qquad j\neq l. \tag{16}$$

Anderson(1999a), 给出了 $\hat{\boldsymbol{\alpha}}_j$ 和 $\hat{\gamma}_j$, $\hat{\boldsymbol{\alpha}}_l$ 的渐近协方差. 注意到 h_{ij}^* 线性依赖于 $(u_{i\alpha},v_{i\alpha})$, $(u_{i\alpha},v_{i\alpha})$ 和 $(u_{j\alpha},v_{j\alpha}),i\neq j$, 是不相关的. 协方差 (14) 不依赖于 $(\boldsymbol{U},\boldsymbol{V})$ 是正态的.

现在假设 $\boldsymbol{\Gamma}_{12}$ 的秩是 $k<p_1$. 定义 $\boldsymbol{H}_1=(\boldsymbol{H}_{11}',\boldsymbol{H}_{21}')'$ 是满足 (4) 式的 $\boldsymbol{H}$ 的前 k 列, 定义 $\hat{\boldsymbol{\Lambda}}_1=\mathrm{diag}(\hat{\boldsymbol{\lambda}}_1,\cdots,\hat{\boldsymbol{\lambda}}_k)$. 则 $\boldsymbol{H}_1$ 满足 (6) 式, $\boldsymbol{H}_1^*$ 满足 (8), (9), (10) 和 (11). 则 λ_i^* 是由 (12) 式给出的, $i=1,\cdots,k$.

(4) 式的后 p_1-k 个列是

$$\begin{bmatrix}\boldsymbol{\Lambda}_1\boldsymbol{S}_{UV}^{*12}\boldsymbol{C}_2\\\boldsymbol{0}\end{bmatrix}=\begin{bmatrix}-\boldsymbol{\Lambda}_1^2\boldsymbol{H}_{12}^*\\\boldsymbol{0}\end{bmatrix}+o_p(1). \tag{17}$$

因此

$$\boldsymbol{\Lambda}_1\boldsymbol{H}_{12}^*=-\boldsymbol{S}_{UV}^{*12}\boldsymbol{C}_2+o_p(1)=-\boldsymbol{S}_{WV}^{*12}\boldsymbol{C}_2+o_p(1). \tag{18}$$

13.7.2 一组变量是随机的，另外一组变量是非随机的

现在我们考虑 (2) 中的 $\boldsymbol{X}^{(2)}$ 是非随机的情况，其中 $\mathrm{E}(\boldsymbol{Z}_\alpha)=\boldsymbol{0}$, $\mathrm{E}(\boldsymbol{Z}_\alpha\boldsymbol{Z}'_\alpha)=\boldsymbol{\Sigma}_{ZZ}$. 我们观察到 $\boldsymbol{X}=\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$. 假设

$$\boldsymbol{S}_{22}=\frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{x}_\alpha^{(2)}\boldsymbol{x}_\alpha^{(2)\prime}\to\boldsymbol{\Sigma}_{22},\tag{19}$$

$\boldsymbol{\Sigma}_{22}$ 是非奇异的. 则

$$\boldsymbol{S}_{11}=\frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{x}_\alpha^{(1)}\boldsymbol{x}_\alpha^{(1)\prime}=\boldsymbol{\beta}\boldsymbol{S}_{22}\boldsymbol{\beta}+\boldsymbol{S}_{Z2}\boldsymbol{\beta}'+\boldsymbol{\beta}\boldsymbol{S}_{2Z}+\boldsymbol{S}_{ZZ}\xrightarrow{p}\boldsymbol{\beta}\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}'+\boldsymbol{\Sigma}_{ZZ},\tag{20}$$

$$\boldsymbol{S}_{12}=\frac{1}{n}\sum_{\alpha=1}^{n}\boldsymbol{x}_\alpha^{(1)}\boldsymbol{x}_\alpha^{(2)\prime}=\boldsymbol{\beta}\boldsymbol{S}_{22}+\boldsymbol{S}_{Z2}\xrightarrow{p}\boldsymbol{\beta}\boldsymbol{\Sigma}_{22}.\tag{21}$$

定义 $\lambda,\boldsymbol{\alpha}$, 和 $\boldsymbol{\gamma}$ 是

$$\begin{bmatrix}-\lambda(\boldsymbol{\Sigma}_{ZZ}+\boldsymbol{\beta}\boldsymbol{S}_{22}\boldsymbol{\beta}') & \boldsymbol{\beta}\boldsymbol{S}_{22}\\ \boldsymbol{S}_{22}\boldsymbol{\beta} & -\lambda\boldsymbol{S}_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\gamma}\end{bmatrix}=\boldsymbol{0},\tag{22}$$

$$\boldsymbol{\alpha}'(\boldsymbol{\Sigma}_{ZZ}+\boldsymbol{\beta}\boldsymbol{S}_{22}\boldsymbol{\beta}')\boldsymbol{\alpha}=1,\qquad \boldsymbol{\gamma}'\boldsymbol{S}_{22}\boldsymbol{\gamma}-1\tag{23}$$

的解.

我们首先假设 $p_1=p_2,\lambda_1>\cdots>\lambda_{p_1}>0$. 则 (22) 和 (23) 还有 $\alpha_{ii}>0$ 定义

$$\mathrm{diag}(\lambda_1,\cdots,\lambda_{p_1})=\boldsymbol{\Lambda}_n,\qquad(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{p_1})=\boldsymbol{A}_n,\qquad(\boldsymbol{\gamma}_1,\cdots,\boldsymbol{\gamma}_{p_1})=\boldsymbol{\Gamma}_n.\tag{24}$$

令 $\boldsymbol{U}=\boldsymbol{A}'_n\boldsymbol{X}^{(1)}$, 那么 $\boldsymbol{v}_\alpha=\boldsymbol{\Gamma}'_n\boldsymbol{x}_\alpha$, $\alpha=1,\cdots,n$, $\boldsymbol{W}=\boldsymbol{A}'_n\boldsymbol{Z}$, $\boldsymbol{\Theta}=\boldsymbol{A}'_n\boldsymbol{\beta}(\boldsymbol{\Gamma}'_n)^{-1}=\boldsymbol{\Lambda}$, $\boldsymbol{H}=\boldsymbol{\Gamma}_n^{-1}\hat{\boldsymbol{\Gamma}}$. 则 $\boldsymbol{H}$ 和 $\hat{\boldsymbol{\Gamma}}$ 满足 (4). 那么 $\boldsymbol{S}_{VV}=\boldsymbol{I}$,

$$\boldsymbol{S}_{UV}=\boldsymbol{\Theta}\boldsymbol{S}_{VV}+\boldsymbol{S}_{WV}=\boldsymbol{\Theta}+\boldsymbol{S}_{WV}\xrightarrow{p}\boldsymbol{\Theta},\tag{25}$$

$$\boldsymbol{S}_{UU}=\boldsymbol{\Theta}\boldsymbol{S}_{VV}\boldsymbol{\Theta}+\boldsymbol{\Theta}\boldsymbol{S}_{VW}+\boldsymbol{S}_{WV}\boldsymbol{\Theta}+\boldsymbol{S}_{WW}\xrightarrow{p}\boldsymbol{I}.\tag{26}$$

则 (4) 式可以写成

$$(\boldsymbol{\Lambda}+\boldsymbol{S}_{VW})(\boldsymbol{\Lambda}^2+\boldsymbol{\Lambda}\boldsymbol{S}_{VW}+\boldsymbol{S}_{WV}\boldsymbol{\Lambda}+\boldsymbol{S}_{WW})^{-1}(\boldsymbol{\Lambda}+\boldsymbol{S}_{WV})\boldsymbol{H}=\boldsymbol{H}\hat{\boldsymbol{\Lambda}}^2.\tag{27}$$

注意到 $\boldsymbol{S}_{VW}\xrightarrow{p}\boldsymbol{0},\boldsymbol{S}_{WW}\xrightarrow{p}\boldsymbol{I}-\boldsymbol{\Lambda}^2$, 因此 $\boldsymbol{H}\xrightarrow{p}\boldsymbol{I},\hat{\boldsymbol{\Lambda}}\xrightarrow{p}\boldsymbol{\Lambda}$.

令 $\boldsymbol{S}^*_{VW}=\sqrt{n}\boldsymbol{S}_{VW},\boldsymbol{S}^*_{WW}=\sqrt{n}[\boldsymbol{S}_{WW}-(\boldsymbol{I}-\boldsymbol{\Lambda}^2)]$. 则由 (27) 式得到

$$\begin{aligned}&(\boldsymbol{I}-\boldsymbol{\Lambda}^2)\boldsymbol{S}^*_{VW}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}\boldsymbol{S}^*_{WV}(\boldsymbol{I}-\boldsymbol{\Lambda}^2)-\boldsymbol{\Lambda}\boldsymbol{S}^*_{WW}\boldsymbol{\Lambda}+\boldsymbol{\Lambda}^2\boldsymbol{H}^*\\&=\boldsymbol{H}^*\boldsymbol{\Lambda}^2+2\boldsymbol{\Lambda}\boldsymbol{\Lambda}^*+o_p(1).\end{aligned}\tag{28}$$

(28) 式的对角项是

$$\lambda_i^*=(1-\lambda_i^2)\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}v_{i\alpha}w_{i\alpha}-\frac{1}{2}\lambda_i\frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}[w_{i\alpha}^2-(1-\lambda_i^2)]+o_p(1).\tag{29}$$

因为在 $\boldsymbol{W}$ 是正态分布的假设下,

$$\mathrm{E}(v_{i\alpha}w_{i\alpha})^2 = v_{i\alpha}^2(1-\lambda_i^2), \tag{30}$$

$$\mathrm{E}[w_{i\alpha}^2-(1-\lambda_i^2)]^2 = 2(1-\lambda_i^2)^2, \tag{31}$$

$\sqrt{n}(\hat{\lambda}_i-\lambda_i)$ 的极限分布是 $N[0,(1-\lambda_i^2)^2(1-\frac{1}{2}\lambda_i^2)]$. 注意到这个方差比 $\boldsymbol{X}^{(2)}$ 是随机的情况下的小.

从 (28) 式我们得到

$$(\lambda_j^2-\lambda_i^2)h_{ij}^* = \frac{1}{\sqrt{n}}\sum_{\alpha=1}^{n}\big[(1-\lambda_i^2)v_{i\alpha}w_{j\alpha}\lambda_j + \lambda_i w_{i\alpha}v_{j\alpha}(1-\lambda_j^2) - \lambda_j w_{i\alpha}w_{j\alpha}\lambda_j\big] + o_p(1). \tag{32}$$

则有

$$(\lambda_j^2-\lambda_i^2)^2\mathrm{E}(h_{ij}^*)^2 \to (1-\lambda_i^2)(1-\lambda_j^2)(\lambda_i^2-\lambda_j^2-\lambda_i^2\lambda_j^2). \tag{33}$$

方程 $\boldsymbol{H}'\boldsymbol{S}_{VV}\boldsymbol{H}=\boldsymbol{I}$ 包含了 $\boldsymbol{H}'\boldsymbol{H}=\boldsymbol{I}$, 导出了 $\boldsymbol{H}^*=-\boldsymbol{H}^{*'}+o_p(1)$, 即 $h_{ij}^*=-h_{ji}^*+o_p(1)$.

现在假设 $\boldsymbol{\beta}$ 的秩是 $k<p_1=p_2$. 则有 $\boldsymbol{\Lambda}=\mathrm{diag}(\boldsymbol{\Lambda}_1,\boldsymbol{0})$, 其中 $\boldsymbol{\Lambda}_1=\mathrm{diag}(\lambda_1,\cdots,\lambda_k)$. 令 $\boldsymbol{\Gamma}=(\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2)$, 其中 $\boldsymbol{\Gamma}_1$ 有 k 列, $\boldsymbol{\Gamma}_2$ 有 p_1-k 列. 定义划分 (5) 为 k 和 p_1-k 行和列部分. (4) 式的概率极限暗示了 $\boldsymbol{H}_{11}\xrightarrow{p}\boldsymbol{I}_k, \boldsymbol{H}_{12}\xrightarrow{p}\boldsymbol{0}, \boldsymbol{H}_{21}\xrightarrow{p}\boldsymbol{0}, \boldsymbol{H}_{22}'\boldsymbol{H}_{22}\xrightarrow{p}\boldsymbol{I}$. 令 $\boldsymbol{H}_{22}$ 的奇异值分解是 $\boldsymbol{EJF}$, 其中 $\boldsymbol{J}$ 是对角矩阵, 阶是 p_1-k, $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是正交矩阵, 阶为 p_1-k. 定义 $\boldsymbol{C}_2=\boldsymbol{EF}$. 用 $\boldsymbol{S}_{VW}^*=\sqrt{n}(\boldsymbol{S}_{VW}-\boldsymbol{\Lambda}), \boldsymbol{S}_{WV}^*=\sqrt{n}(\boldsymbol{S}_{UV}^*-\boldsymbol{\Lambda}), \boldsymbol{S}_{WW}^*=\sqrt{n}[\boldsymbol{S}_{WW}-(\boldsymbol{I}-\boldsymbol{\Lambda}^2)], \boldsymbol{H}_{11}^*=\sqrt{n}(\boldsymbol{H}_{11}-\boldsymbol{I}), \boldsymbol{H}_{12}^*=\sqrt{n}\boldsymbol{H}_{12}, \boldsymbol{H}_{21}^*=\sqrt{n}\boldsymbol{H}_{21}, \boldsymbol{H}_{22}^*=\sqrt{n}(\boldsymbol{H}_{22}-\boldsymbol{C}_2)=\sqrt{n}\boldsymbol{E}(\boldsymbol{J}-\boldsymbol{I})\boldsymbol{F}$, 把 (4) 展开就得到

$$\begin{bmatrix} \boldsymbol{\Lambda}_1\boldsymbol{S}_{WV}^{*11}(\boldsymbol{I}-\boldsymbol{\Lambda}_1^2)+(\boldsymbol{I}-\boldsymbol{\Lambda}_1^2)\boldsymbol{S}_{VW}^{*11}\boldsymbol{\Lambda}_1-\boldsymbol{\Lambda}_1\boldsymbol{S}_{WV}^{*11}\boldsymbol{\Lambda}_1 & \boldsymbol{\Lambda}_1\boldsymbol{S}_{WV}^{*12}\boldsymbol{C}_2 \\ \boldsymbol{S}_{VW}^{*21}\boldsymbol{\Lambda}_1 & \boldsymbol{0} \end{bmatrix} \tag{34}$$
$$= \begin{bmatrix} 2\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1^*+\boldsymbol{H}_{11}^*\boldsymbol{\Lambda}_1^2-\boldsymbol{\Lambda}_1^2\boldsymbol{H}_{11}^* & -\boldsymbol{\Lambda}_1^2\boldsymbol{H}_{12}^* \\ \boldsymbol{H}_{21}^*\boldsymbol{\Lambda}_1^2 & \boldsymbol{0} \end{bmatrix} + o_p(1).$$

(34) 式的第 i 个对角项是 (29), $i=1,\cdots,k$. 左上三角子矩阵的 (i,j) 元是 (32) 式, 对 $i\neq j$, $i,j=1,\cdots,k$. (34) 的另外两个子矩阵方程是

$$\boldsymbol{\Lambda}_1\boldsymbol{H}_{12}^* = -\boldsymbol{S}_{WV}^{*12}\boldsymbol{C}_2+o_p(1), \tag{35}$$

$$\boldsymbol{H}_{21}^*\boldsymbol{\Lambda}_1 = \boldsymbol{S}_{VW}^{*21}+o_p(1). \tag{36}$$

方程 $\boldsymbol{I}=\boldsymbol{H}'\boldsymbol{S}_{VV}\boldsymbol{H}=\boldsymbol{H}'\boldsymbol{H}$ 导出

$$\begin{bmatrix} \boldsymbol{H}_{11}^*+\boldsymbol{H}_{11}^* & \boldsymbol{H}_{12}^*+\boldsymbol{H}_{21}^{*'}\boldsymbol{C}_2 \\ \boldsymbol{C}_2'\boldsymbol{H}_{21}^*+\boldsymbol{H}_{12}^{*'} & \boldsymbol{C}_2\boldsymbol{H}_{22}^*+\boldsymbol{H}_{22}^{*'}\boldsymbol{C}_2 \end{bmatrix} = \boldsymbol{0}+o_p(1). \tag{37}$$

(37) 的非对角子矩阵同 (35) 和 (36) 一致.

13.7.3 降秩回归估计量

当 $\boldsymbol{\beta}$ 的秩给定是 $k(<p_1)$ 时, $\boldsymbol{\beta}$ 的极大似然估计量是

$$\hat{\boldsymbol{B}}_k = \boldsymbol{S}_{12}\hat{\boldsymbol{\Gamma}}_1\hat{\boldsymbol{\Gamma}}_1'. \tag{38}$$

见 12.7 节. 由 (3) 式 $\boldsymbol{\Theta}$ 降秩回归估计量是

$$\hat{\boldsymbol{\Theta}}_k = \boldsymbol{S}_{UV}\boldsymbol{H}_1\boldsymbol{H}_1'. \tag{39}$$

假设 $\boldsymbol{X}^{(2)}$ 是随机的, $\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\Theta}_1, \boldsymbol{0}) = \mathrm{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{0})$. 我们定义 $\boldsymbol{\Theta}_k^* = \sqrt{n}(\hat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}), \boldsymbol{H}_1^* = \sqrt{n}(\boldsymbol{H}_1 - \boldsymbol{I}_{(k)}), \boldsymbol{S}_{UV}^* - \sqrt{n}(\boldsymbol{S}_{UV} - \boldsymbol{\Lambda}), \boldsymbol{S}_{VV}^* = \sqrt{n}(\boldsymbol{S}_{VV} - \boldsymbol{I})$. 从 $\boldsymbol{H}_1'\boldsymbol{S}_{VV}\boldsymbol{H}_1 = \boldsymbol{I}$ 我们得到 $\boldsymbol{H}_{11}^* + \boldsymbol{H}_{11}^{*'} = -\boldsymbol{S}_{VV}^{*11} + o_p(1)$. 从 (39) 和 (9) 我们得到

$$\begin{aligned}\hat{\boldsymbol{\Theta}}_k^* &= \begin{bmatrix} \boldsymbol{S}_{UV}^{*11} + \boldsymbol{\Lambda}_1(\boldsymbol{H}_{11}^* + \boldsymbol{H}_{11}^{*'}) & \boldsymbol{\Lambda}_1\boldsymbol{H}_{21}^{*'} \\ \boldsymbol{S}_{UV}^{*21} & \boldsymbol{0} \end{bmatrix} + o_p(1) \\ &= \begin{bmatrix} \boldsymbol{S}_{WV}^{*11} & \boldsymbol{S}_{WV}^{*12} \\ \boldsymbol{S}_{WV}^{*21} & \boldsymbol{0} \end{bmatrix} + o_p(1).\end{aligned} \tag{40}$$

我们由秩条件 $\hat{\boldsymbol{\Theta}} = \boldsymbol{S}_{UV}\boldsymbol{S}_{VV}^{-1}$, 比较 $\hat{\boldsymbol{\Theta}}_k$ 和无约束的极大似然估计量. 则

$$\begin{aligned}\hat{\boldsymbol{\Theta}}^* &= \sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) = (\boldsymbol{S}_{UV} - \boldsymbol{\Theta}\boldsymbol{S}_{VV})\boldsymbol{S}_{VV}^{-1} \\ &= \boldsymbol{S}_{WV}^* + o_p(1) = \begin{bmatrix} \boldsymbol{S}_{WV}^{*11} & \boldsymbol{S}_{WV}^{*12} \\ \boldsymbol{S}_{WV}^{*21} & \boldsymbol{S}_{WV}^{*22} \end{bmatrix} + o_p(1),\end{aligned} \tag{41}$$

由于 $\boldsymbol{S}_{VV} \xrightarrow{p} \boldsymbol{I}$. 秩约束的效应是用 $\boldsymbol{0}$(参数值) 代替 $\boldsymbol{S}_{WV}^*$ 的低阶右下子矩阵.

因为 $\boldsymbol{S}_{WV}^* = (1/\sqrt{n})\sum_{\alpha=1}^n \boldsymbol{W}_\alpha\boldsymbol{V}_\alpha'$, 我们有 $\mathrm{vec}\boldsymbol{S}_{WV}^* = (1/\sqrt{n})\sum_{\alpha=1}^n(\boldsymbol{V}_\alpha \otimes \boldsymbol{W}_\alpha)$. 由于 $\boldsymbol{V}_\alpha$ 和 $\boldsymbol{W}_\alpha$ 是独立的, 有

$$\begin{aligned}\mathrm{E}(\mathrm{vec}\boldsymbol{S}_{WV}^*(\mathrm{vec}\boldsymbol{S}_{WV}^*)') &= \mathrm{E}(\boldsymbol{V}\boldsymbol{V}') \otimes \mathrm{E}(\boldsymbol{W}\boldsymbol{W}') = \boldsymbol{I} \otimes (\boldsymbol{I} - \boldsymbol{\Lambda}^2) \\ &= \mathrm{diag}(\boldsymbol{I} - \boldsymbol{\Lambda}^2, \cdots, \boldsymbol{I} - \boldsymbol{\Lambda}^2),\end{aligned} \tag{42}$$

其中 $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{0}), \boldsymbol{I} - \boldsymbol{\Lambda}^2 = \mathrm{diag}(\boldsymbol{I} - \boldsymbol{\Lambda}_1^2, \boldsymbol{I})$. 另一方面

$$\begin{aligned}\mathrm{vec}\hat{\boldsymbol{\Theta}}_k^* &= \mathrm{vec}\frac{1}{\sqrt{n}}\sum_{\alpha=1}^n \left[\boldsymbol{W}_\alpha\boldsymbol{V}_\alpha^{(1)'}, \begin{pmatrix} \boldsymbol{W}_\alpha^{(1)} \\ \boldsymbol{0} \end{pmatrix}\boldsymbol{V}_\alpha^{(2)'}\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{\alpha=1}^n \begin{bmatrix} \boldsymbol{V}_\alpha^{(1)} \otimes \boldsymbol{W}_\alpha \\ \boldsymbol{V}_\alpha^{(2)} \otimes \begin{pmatrix} \boldsymbol{W}_\alpha^{(1)} \\ \boldsymbol{0} \end{pmatrix} \end{bmatrix} + o_p(1),\end{aligned} \tag{43}$$

其中 $\boldsymbol{V}_\alpha = (\boldsymbol{V}_\alpha^{(1)'}, \ \boldsymbol{V}_\alpha^{(2)'})', \boldsymbol{W}_\alpha = (\boldsymbol{W}_\alpha^{(1)'}, \boldsymbol{W}_\alpha^{(2)'})'$. 则

$$\begin{aligned}\mathrm{E}(\mathrm{vec}\hat{\boldsymbol{\Theta}}_k^*(\mathrm{vec}\hat{\boldsymbol{\Theta}}_k^*)') &\to \begin{bmatrix} \boldsymbol{I}_k \otimes (\boldsymbol{I}_{p_1} - \boldsymbol{\Lambda}^2) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{p_2-k} \otimes \begin{bmatrix} \boldsymbol{I}_k - \boldsymbol{\Lambda}_1^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \end{bmatrix} \\ &= \mathrm{diag}(\boldsymbol{I}_{p_1} - \boldsymbol{\Lambda}^2, \cdots, \boldsymbol{I}_{p_1} - \boldsymbol{\Lambda}^2, \boldsymbol{I}_k - \boldsymbol{\Lambda}_1^2, \boldsymbol{0}, \cdots, \boldsymbol{I}_k - \boldsymbol{\Lambda}_1^2, \boldsymbol{0}),\end{aligned} \tag{44}$$

其中 $I_{p_1}-\Lambda^2$ 有 k 块, $\mathrm{diag}(I_k-\Lambda_1^2,0)$ 有 p_1-k 块.

在原始坐标系下

$$\begin{aligned}\mathrm{vec}(\hat{B}_k-\boldsymbol{\beta})&=\mathrm{vec}\left[(A')^{-1}(\hat{\Theta}_k-\Theta)\Gamma'\right]\\&=\left[\Gamma\otimes(A')^{-1}\right]\mathrm{vec}(\hat{\Theta}-\Theta)\\&=\left[(\Gamma_1,\Gamma_2)\otimes\Sigma_{ZZ}(A_1,A_2)(I-\Lambda^2)^{-1}\right]\mathrm{vec}(\hat{\Theta}-\Theta).\end{aligned}\tag{45}$$

从 (44) 和 (45) 我们得到

$$\begin{aligned}&\mathrm{E}\left(\mathrm{vec}\, n(\hat{B}_k-\boldsymbol{\beta})[\mathrm{vec}(\hat{B}_k-\boldsymbol{\beta})]'\right)\\&\to[(\Gamma_1\Gamma_1')\otimes\Sigma_{ZZ}A(I_p-\Lambda^2)^{-1}]+[\Gamma_2\Gamma_2'\otimes\Sigma_{ZZ}A_1(I_k-\Lambda_1^2)^{-1}A_1'\Sigma_{ZZ}]\\&=[\Gamma_1\Gamma_1'\otimes\Sigma_{ZZ}]+[\Gamma_2\Gamma_2'\otimes\Sigma_{ZZ}A_1(I_k-\Lambda_1^2)^{-1}A_1'\Sigma_{ZZ}]\\&=\Sigma_{XX}^{-1}\otimes\Sigma_{ZZ}-(\Gamma_2\Gamma_2'\otimes\Sigma_{ZZ}A_2A_2'\Sigma_{ZZ}).\end{aligned}\tag{46}$$

如果我们定义 $\Omega=\Sigma_{YX}\Gamma_1=\Sigma_{ZZ}A_1\Lambda_1(I-\Lambda_1^2)^{-1}$ 和 $\Pi=\Gamma_1$, 那么 $\boldsymbol{\beta}=\Omega\Pi'$. 我们有

$$\Omega(\Omega'\Sigma_{ZZ}^{-1}\Omega)^{-1}\Omega'=\Sigma_{ZZ}-\Sigma_{ZZ}A_2A_2'\Sigma_{ZZ},\tag{47}$$

$$\Pi(\Pi'\Sigma_{XX}\Pi)^{-1}\Pi'=\Gamma_1\Gamma_1'=\Sigma_{XX}^{-1}-\Gamma_2\Gamma_2'.\tag{48}$$

因此 (46) 式可以写成

$$\begin{aligned}\mathrm{E}\left(\mathrm{vec}\,\hat{B}_k^*(\mathrm{vec}\hat{B}_k^*)'\right)\to\Sigma_{XX}^{-1}\otimes\Sigma_{ZZ}-[\Sigma_{XX}^{-1}-\Pi(\Pi'\Sigma_{XX}\Pi)^{-1}\Pi']\\\otimes[\Sigma_{ZZ}-\Omega(\Omega'\Sigma_{ZZ}^{-1}\Omega)^{-1}\Omega'].\end{aligned}\tag{49}$$

定理 13.7.1 令 $(X^{(1)'},X^{(2)'})'(\alpha=1,\cdots,n)$ 是随机向量 x_α 的观测, x_α 有均值 0 和协方差矩阵 Σ. 令 $\boldsymbol{\beta}=\Sigma_{12}\Sigma_{22}^{-1}$. 令 $\hat{\Gamma}_1$ 的列满足 (1) 和 $\hat{\gamma}_{ii}>0$. 假设 $X^{(1)}-\boldsymbol{\beta}X^{(2)}=Z$ 是和 $X^{(2)}$ 独立的. 当 $\hat{B}_k=S_{12}\hat{\Gamma}_1\hat{\Gamma}_1'$ 时, $\mathrm{vec}\hat{B}_k^*=\sqrt{n}\mathrm{vec}(\hat{B}_k-\boldsymbol{\beta})$ 的极限分布是正态的, 并且均值为 0, 协方差阵为 (46) 或者 (49).

注意到对任意的非奇异的 M, 有 $\boldsymbol{\beta}=\Omega\Pi'=\Omega M'(\Pi M^{-1})'$; (47) 和 (48) 式关于变换 $\Omega\to\Omega M$ 和 $\Pi\to\Pi M^{-1}$ 是不变的. 因此 (49) 式对任意的因子分解 $\boldsymbol{\beta}=\Omega\Pi'$ 成立.

$\hat{B}_k^*$ 的极限分布仅仅取决于 $\sqrt{n}S_{Z2}S_{22}^{-1}=(A')^{-1}S_{WV}^*S_{VV}^{-1}\Gamma'$, 因此最小二乘估计量 $\hat{B}$ 的渐近正态性在相同的条件下成立.

现在我们假设 $X_\alpha^{(2)}=x_\alpha^{(2)}(\alpha=1,\cdots,n)$ 是非随机的并且 (19) 成立. 模型是 (2); 在变换坐标系下, $[U=A_n'x^{(1)},v_\alpha=\Gamma_n'x_\alpha^{(2)},W=A_n'Z,\Theta=A_n'\boldsymbol{\beta}(\Gamma_n')^{-1}=\Lambda]$ 模型是 (3) 式. $H_1=\Gamma^{-1}\Gamma_1$ 满足 (34) 和 (37). 同样 (39) 式成立, 进而, 当 $V_\alpha=v_\alpha$ 非随机时 (42) 式和 (43) 式成立.

推论 13.7.1 令 $x_1^{(2)},\cdots,x_n^{(2)}$ 是一组向量, 使得 (19) 成立. 令 $x_\alpha^{(1)}=\boldsymbol{\beta}x_\alpha^{(2)}+z_\alpha,\alpha=1,\cdots,n$, 其中 z_α 是随机向量 Z 的一个观测, 有 $\mathrm{E}(Z)=0$ 和 $\mathrm{E}(ZZ')=\Sigma_{ZZ}$. 假设 $\boldsymbol{\beta}$ 有秩 k. 则 $\sqrt{n}\mathrm{vec}(\hat{B}_k-\boldsymbol{\beta})$ 的极限分布是正态的, 有均值 0, 协方差 (46) 或 (49).

13.8 椭球等高分布

13.8.1 椭球等高观测

令 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ 是一个随机向量 $\boldsymbol{X}$ 的 N 个观测, $\boldsymbol{X}$ 有密度

$$|\boldsymbol{\Psi}|^{-\frac{1}{2}}g\left[(\boldsymbol{x}-\boldsymbol{v})'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{v})\right], \tag{1}$$

其中 $\boldsymbol{\Psi}$ 是正定矩阵, $R^2=(\boldsymbol{x}-\boldsymbol{v})'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}-\boldsymbol{v})$, $\mathrm{E}(R^2)<\infty$. 定义 $\kappa=p\mathrm{E}(R^4)/[(\mathrm{E}(R^2))^2(p+2)]-1$. 则 $\mathrm{E}(\boldsymbol{X})=\boldsymbol{v}=\mu, \mathrm{E}((\boldsymbol{X}-\boldsymbol{v})(\boldsymbol{X}-\boldsymbol{v})')=(\mathrm{E}(R^2)/p)\boldsymbol{\Psi}=\boldsymbol{\Sigma}$. 定义 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 是样本均值和协方差阵. 由下式定义正交矩阵 $\boldsymbol{\beta}$ 和 $\boldsymbol{B}$ 以及对角矩阵 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{L}$,

$$\boldsymbol{\Sigma}=\boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}', \qquad \boldsymbol{S}=\boldsymbol{B}\boldsymbol{L}\boldsymbol{B}', \tag{2}$$

$\lambda_1>\cdots>\lambda_p, l_1>\cdots>l_p, \beta_{i1}\geqslant 0, b_{i1}\geqslant 0, i=1,\cdots,p$. 如 13.5.1 节, 定义 $\boldsymbol{T}=\boldsymbol{\beta}'\boldsymbol{S}\boldsymbol{\beta}=\boldsymbol{Y}\boldsymbol{L}\boldsymbol{Y}'$, 其中 $\boldsymbol{Y}=\boldsymbol{\beta}'\boldsymbol{B}$ 是正交的, $y_{i1}\geqslant 0$. 则有 $\mathrm{E}(\boldsymbol{T})=\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}=\boldsymbol{\Lambda}$.

$\sqrt{N}\mathrm{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})$ 和 $\sqrt{N}\mathrm{vec}(\boldsymbol{T}-\boldsymbol{\Lambda})$ 的极限协方差是

$$\begin{aligned}&\lim_{N\to\infty}N\mathrm{E}(\mathrm{vec}\,(\boldsymbol{S}-\boldsymbol{\Sigma})[\mathrm{vec}\,(\boldsymbol{S}-\boldsymbol{\Sigma})]')\\ =&(\kappa+1)(\boldsymbol{I}_{p^2}+K_{pp})(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\kappa\,\mathrm{vec}\,\boldsymbol{\Sigma}(\mathrm{vec}\,\boldsymbol{\Sigma})',\end{aligned} \tag{3}$$

$$\begin{aligned}&\lim_{N\to\infty}N\mathrm{E}(\mathrm{vec}\,(\boldsymbol{T}-\boldsymbol{\Lambda})[\mathrm{vec}\,(\boldsymbol{T}-\boldsymbol{\Lambda})]')\\ =&(\kappa+1)(\boldsymbol{I}_{p^2}+\boldsymbol{K}_{pp})+\kappa\,\mathrm{vec}\,\boldsymbol{I}_p(\mathrm{vec}\,\boldsymbol{I}_p)'.\end{aligned} \tag{4}$$

用分量的方法, $\mathrm{E}(t_{ij})=\lambda_i\delta_{ij}$ 和

$$\begin{aligned}&\lim_{N\to\infty}N\,\mathrm{E}((t_{ij}-\lambda_i\delta_{ij})(t_{kl}-\lambda_k\delta_{kl}))\\ =&(\kappa+1)(\lambda_i\lambda_j\delta_{ik}\delta_{jl}+\lambda_i\lambda_k\delta_{il}\delta_{jk})+\kappa\lambda_i\lambda_k\delta_{ij}\delta_{kl}.\end{aligned} \tag{5}$$

令 $\sqrt{N}(\boldsymbol{T}-\boldsymbol{\Lambda})=\boldsymbol{U}, \sqrt{N}(\boldsymbol{L}-\boldsymbol{\Lambda})=\boldsymbol{D}, \sqrt{N}(\boldsymbol{Y}-\boldsymbol{I}_p)=\boldsymbol{W}$. 集合 $u_{11},\cdots,u_{pp}$ 是和集合 $(u_{12},\cdots,u_{p-1,p})$ 渐近独立的; 协方差 $u_{ij}(i\neq j)$ 是和方差 $(\kappa+1)\lambda_i\lambda_j$ 相互独立的; $u_{ii}=d_i$ 的方差收敛到 $(3\kappa+2)\lambda_i^2$; $u_{ii}=d_i$ 和 $u_{kk}=d_k(i\neq k)$ 的协方差收敛到 $\kappa\lambda_i\lambda_k$. $w_{ij}(i\neq j)$ 的极限分布是 $u_{ij}/(\lambda_j-\lambda_i)$ 的极限分布. 因此 $w_{ij}(i<j)$ 是渐近相互独立的, 并且 $\mathrm{E}(w_{ij}^2)=(\kappa+1)\lambda_i\lambda_j/(\lambda_j-\lambda_i)^2$. 这些方差和协方差对 $\kappa=0$ 在正态情况下成立.

定理 13.8.1 由 (2) 定义对角矩阵 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{L}$, 正交阵 $\boldsymbol{\beta}$ 和 $\boldsymbol{B}$, $\lambda_1>\cdots>\lambda_p, l_1>\cdots>l_p, \beta_{i1}\geqslant 0, b_{i1}\geqslant 0, i=1,\cdots,p$. 定义 $\boldsymbol{G}=\sqrt{N}(\boldsymbol{B}-\boldsymbol{\beta})$, 对角矩阵 $\boldsymbol{D}=\sqrt{N}(\boldsymbol{L}-\boldsymbol{\Lambda})$. 当 $\boldsymbol{G}$ 和 $\boldsymbol{D}$ 独立时, $\boldsymbol{G}$ 和 $\boldsymbol{D}$ 的极限分布是正态的. d_i 的方差是 $(2+3\kappa)\lambda_i^2$, d_i 和 d_k 的协方差是 $\kappa\lambda_i\lambda_k$. $\boldsymbol{g}_i$ 的协方差是

$$\mathrm{ACov}(\boldsymbol{g}_i)=(1+\kappa)\sum_{\substack{k=1\\k\neq i}}^{p}\frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\boldsymbol{\beta}_k\boldsymbol{\beta}_k', \tag{6}$$

$\boldsymbol{g}_i$ 和 $\boldsymbol{g}_j$ 的协方差阵是

$$\mathrm{ACov}(\boldsymbol{g}_i, \boldsymbol{g}_j) = -(1+\kappa)\frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_i', \quad i \neq j. \tag{7}$$

证明 证明和定理 13.5.1 的相同, 除了用 (4) 式代替 $\kappa = 0$ 时的 (4) 式. ■

在 11.7.3 节, 当 q 个较小的总体根相等时, 我们用 q 个样本根的渐近分布. 令 $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\Lambda}_1, \lambda^*\boldsymbol{I}_q)$, 其中 (对角的)$\boldsymbol{\Lambda}_1$ 的对角元是不同的, 大于 λ^*. 类似以前, 令 $\boldsymbol{U} = \sqrt{N}(\boldsymbol{T}-\boldsymbol{\Lambda})$, 令 $\boldsymbol{U}_{22}$ 是 $\boldsymbol{U}$ 的右下角 $q \times q$ 子矩阵. 令 $\boldsymbol{D}_2$ 和 $\boldsymbol{Y}_{22}$ 是 $\boldsymbol{D}$ 和 $\boldsymbol{Y}$ 的右下角 $q \times q$ 子矩阵. 13.5.2 节中证明了 $\boldsymbol{U}_{22} = \boldsymbol{Y}_{22}\boldsymbol{D}_2\boldsymbol{Y}_{22}' + o_P(1)$.

检验原假设 $\lambda_{p-q+1} = \cdots = \lambda_p$ 的准则是

$$\frac{\prod_{i=p-q+1}^{p} l_i}{\left(\sum_{i=p-q+1}^{p} l_i/q\right)^q}. \tag{8}$$

在 11.7.3 节证明了 $-N$ 乘以 (8) 式的对数后有极限分布

$$\begin{aligned} &\frac{1}{2\lambda^{*2}}\left[\mathrm{tr}\boldsymbol{U}_{22}\boldsymbol{U}_{22}' - \frac{1}{q}(\mathrm{tr}\boldsymbol{U}_{22})^2\right] \\ &= \frac{1}{2\lambda^{*2}}\left[2\sum_{\substack{i=p-q+1\\ i<j}}^{p} u_{ij}^2 + \sum_{i=p-q+1}^{p} u_{ii}^2 - \frac{1}{q}\left(\sum_{i=p-q+1}^{p} u_{ii}\right)^2\right]. \end{aligned} \tag{9}$$

项 $\sum_{i<j} u_{ij}^2$ 有极限分布 $(1+\kappa)\lambda^{*2}\chi^2_{q(q-1)/2}$. $(u_{p-q-1,p-q+1}, \cdots, u_{pp})$ 的极限分布是正态的, 有均值 $\mathbf{0}$ 和协方差阵 $\lambda^{*2}[2(1+\kappa)\boldsymbol{I}_q + \kappa\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\lambda^{*2}$. $[\Sigma u_{ii}^2 - (\Sigma u_{ii})^2/q]\lambda^{*2}$ 的极限分布是 $2(1+\kappa)\lambda^{*2}\chi^2_{q-1}$. 因此, (9) 式的极限分布是 $(1+\kappa)\chi^2_{q(q+1)/2-1}$ 的分布.

我们同样对一个协方差阵在另一个协方差阵的度量意义下的特征根和特征向量感兴趣.

定理 13.8.2 令 $\boldsymbol{S}^*$ 是来自 (1) 式的 M 个样本的协方差阵, 令 $\boldsymbol{T}^*$ 是来自 (1) 式中 $\boldsymbol{\Psi}$ 被 $\boldsymbol{\Sigma}$ 取代后的 N 个样本的协方差阵. 令 $\boldsymbol{\Lambda}$ 是对角矩阵, 有对角元 $\lambda_1 > \cdots > \lambda_p (> 0)$, 其中 $\lambda_1, \cdots, \lambda_p$ 是 $|\boldsymbol{\Psi} - \lambda\boldsymbol{\Sigma}| = 0$ 的根. 令 $\boldsymbol{\Gamma} = (\gamma_1, \cdots, \gamma_p)$ 是矩阵, 其中 γ_i 是 $(\boldsymbol{\Psi} - \lambda_i\boldsymbol{\Sigma})\gamma = \mathbf{0}$ 的解, $\gamma'\boldsymbol{\Sigma}\gamma = 1, \gamma_{1i} \geqslant 0$. 令 $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_p^*)$, 对角矩阵 $\boldsymbol{L}^*$ 由

$$(\boldsymbol{S}^* - l\boldsymbol{T}^*)\boldsymbol{x}^* = \mathbf{0} \tag{10}$$

的解构成 $\boldsymbol{x}^{*\prime}\boldsymbol{T}^*\boldsymbol{x}^* = 1, \boldsymbol{x}_1^* \geqslant 0$. 当 $M \to \infty, N \to \infty, M/N \to \eta$ 时, $\boldsymbol{Z}^* = \sqrt{N}(\boldsymbol{X}^* - \boldsymbol{\Gamma})$ 和对角矩阵 $\boldsymbol{D}^* = \sqrt{N}(\boldsymbol{L} - \boldsymbol{\Lambda})$ 的极限分布是正态的, 有下面的协方差:

$$\mathrm{AVar}(d_i) = (2+3\kappa)\lambda_i^2\frac{1+\eta}{\eta}, \tag{11}$$

$$\mathrm{ACov}(d_i, d_j) = \kappa\lambda_i\lambda_j\frac{1+\eta}{\eta}, \tag{12}$$

$$\mathrm{ACov}(z_i) = (1+\kappa)\sum_{\substack{k=1\\ k\neq i}}^{p}\frac{\lambda_i(\lambda_k+\eta\lambda_i)}{(\lambda_k-\lambda_i)^2}\gamma_k\gamma_k' + \frac{2+3\kappa}{4}\gamma_i\gamma_i', \tag{13}$$

$$\mathrm{ACov}(d_i, z_i) = \frac{2+3\kappa}{2}\lambda_i\boldsymbol{\gamma}_i, \tag{14}$$

$$\mathrm{ACov}(z_i^*, z_j^*) = -(1+\kappa)\frac{\lambda_i\lambda_j(1+\eta)}{\eta(\lambda_j-\lambda_i)^2}\boldsymbol{\gamma}_j\boldsymbol{\gamma}_i' + \frac{\kappa}{4}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_j, \qquad i \neq j, \tag{15}$$

$$\mathrm{ACov}(d_i, z_j) = \frac{\kappa}{2}\lambda_i\boldsymbol{\gamma}_j. \tag{16}$$

证明 变换 $\boldsymbol{S}^*$ 和 $\boldsymbol{T}^*$ 到 $\boldsymbol{S} = \boldsymbol{\Gamma}'\boldsymbol{S}^*\boldsymbol{\Gamma}$ 和 $\boldsymbol{T} = \boldsymbol{\Gamma}'\boldsymbol{T}^*\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ 和 $\boldsymbol{\Sigma}$ 到 $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}'\boldsymbol{\Phi}\boldsymbol{\Gamma}$ 和 $\boldsymbol{I} = \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma}$, $\boldsymbol{X}^*$ 到 $\boldsymbol{X} = \boldsymbol{\Gamma}^{-1}\boldsymbol{X}^* = \boldsymbol{G}^{-1}$. 令 $\boldsymbol{D} = \sqrt{N}(\boldsymbol{L}-\boldsymbol{\Lambda}), \boldsymbol{H} = \sqrt{N}(\boldsymbol{G}-\boldsymbol{I}), \boldsymbol{U} = \sqrt{N}(\boldsymbol{S}-\boldsymbol{\Lambda}), \boldsymbol{V} = \sqrt{N}(\boldsymbol{T}-\boldsymbol{I})$. 矩阵 $\boldsymbol{U}$, $\boldsymbol{V}, \boldsymbol{D}, \boldsymbol{H}$ 有极限正态分布; 它们因 13.6 节的 (20), (21) 和 (22) 而相关联. 从这些和极限分布的协方差我们得到 (11)~(16). ■

13.8.2 椭球等高矩阵分布

令 $\boldsymbol{Y}(p\times N)$ 有密度 $g(\mathrm{tr}\,\boldsymbol{YY}')$. 则 $\boldsymbol{A} = \boldsymbol{YY}'$ 有密度 (引理 13.3.1)

$$\frac{\pi^{\frac{1}{2}pN}|\boldsymbol{A}|^{\frac{1}{2}(N-p-1)}}{\Gamma_p(\frac{1}{2}N)}g(\mathrm{tr}\,\boldsymbol{A}). \tag{17}$$

令 $\boldsymbol{A} = \boldsymbol{BLB}'$, 其中 $\boldsymbol{L}$ 是对角矩阵, 有对角元 $l_1 > \cdots > l_p$, $\boldsymbol{B}$ 是正交的, 有 $b_{i1} \geqslant 0$. 因为 $g(\mathrm{tr}\,\boldsymbol{A}) = g(\sum_{i=1}^p l_i)$, $l_1, \cdots, l_p$ 的密度是 (定理 13.3.4)

$$\frac{\pi^{\frac{1}{2}p^2} g(\sum_{i=1}^p l_i)\prod_{i<j}(l_i - l_j)}{\Gamma_p(\frac{1}{2}p)}, \tag{18}$$

矩阵 $\boldsymbol{B}$ 独立地服从条件 Haar 不变分布.

假设 $\boldsymbol{Y}^*(p\times m)$ 和 $\boldsymbol{Z}^*(p\times n)$ 有密度

$$\boldsymbol{\Psi}^{(m+n)/2}g[\mathrm{tr}(\boldsymbol{Y}^{*'}\boldsymbol{\Psi}^{-1}\boldsymbol{Y}^* + \boldsymbol{Z}^{*'}\boldsymbol{\Psi}^{-1}\boldsymbol{Z}^*)] \qquad (m, n > p).$$

令 $\boldsymbol{C}$ 是矩阵, 使得 $\boldsymbol{C\Psi C}' = \boldsymbol{I}$. 则 $\boldsymbol{Y} = \boldsymbol{CY}^*$ 和 $\boldsymbol{Z} = \boldsymbol{CZ}^*$ 有密度 $g[\mathrm{tr}(\boldsymbol{YY}'+\boldsymbol{ZZ}')]$. 令 $\boldsymbol{A}^* = \boldsymbol{Y}^*\boldsymbol{Y}^{*'}, \boldsymbol{B}^* = \boldsymbol{Z}^*\boldsymbol{Z}^{*'}, \boldsymbol{A} = \boldsymbol{YY}', \boldsymbol{B} = \boldsymbol{ZZ}'$. $|\boldsymbol{A}^* - l\boldsymbol{B}^*| = 0$ 的根正是 $|\boldsymbol{A} - l\boldsymbol{B}| = 0$ 的根. 令 $|\boldsymbol{A} - f(\boldsymbol{A}+\boldsymbol{B})| = 0$ 的根是 $f_1 > \cdots > f_p$, 令 $\boldsymbol{F} = \mathrm{diag}(f_1, \cdots, f_p)$. 由 $\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{E}'\boldsymbol{E}$ 和 $\boldsymbol{A} = \boldsymbol{E}'\boldsymbol{FE}$ 定义 $\boldsymbol{E}(p\times p), e_{i1} \geqslant 0, i = 1, \cdots, p$.

定理 13.8.3 矩阵 $\boldsymbol{E}$ 和 $\boldsymbol{F}$ 是独立的. $\boldsymbol{F}$ 的密度是

$$\frac{\pi^{\frac{1}{2}p^2}\boldsymbol{\Gamma}_p[\frac{1}{2}(m+n)]}{\boldsymbol{\Gamma}_p(\frac{1}{2}n)\boldsymbol{\Gamma}_p(\frac{1}{2}m)\boldsymbol{\Gamma}_p(\frac{1}{2}p)}\prod_{i=1}^p f_i^{\frac{1}{2}(m-p-1)}\prod_{i=1}^p(1-f_i)^{\frac{1}{2}(n-p-1)}\prod_{i<j}(f_i - f_j); \tag{19}$$

$\boldsymbol{E}$ 的密度是

$$\frac{2^p\boldsymbol{\Gamma}_p(\frac{1}{2}p)\pi^{\frac{1}{2}p(n+m-p)}}{2^{\frac{1}{2}p(m+n-2)}\pi^{\frac{1}{2}p^2}\boldsymbol{\Gamma}_p[\frac{1}{2}(m+n)]}|\boldsymbol{E}'\boldsymbol{E}|^{\frac{1}{2}p(m+n-p)}g(\mathrm{tr}\boldsymbol{E}'\boldsymbol{E}). \tag{20}$$

13.2 节得到了观测 $\boldsymbol{Y}, \boldsymbol{Z}$ 的密度

$$(2\pi)^{-\frac{1}{2}p(n+m)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y}+\boldsymbol{Z}'\boldsymbol{Z})} = (2\pi)^{-\frac{1}{2}p(n+m)}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})}, \tag{21}$$

13.7 节得到 $g[\mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{Z}'\boldsymbol{Z})] = g[\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B})]$. 根的分布不依赖于 $g(\cdot)$ 的形式; $\boldsymbol{E}$ 的分布仅仅依赖于 $\boldsymbol{E}'\boldsymbol{E} = \boldsymbol{A}+\boldsymbol{B}$. 13.2 节的代数列出更多一般的情况.

习 题

13.1 (13.2 节) 对 $p=2$ 通过直接计算雅可比行列式证明定理 13.2.1.

13.2 (13.2 节) 对 $p=2$ 通过用角的余弦和正弦表示正交矩阵 $\boldsymbol{C}$ 的方法来直接证明定理 13.3.2.

13.3 (13.2 节) 当 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是二阶, 并且各自服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)$ 时, 考虑 $|\boldsymbol{A}-l\boldsymbol{B}|=0$ 的根的分布.

(a) 求最大的根的分布.

(b) 求最小的根的分布.

(c) 求根的和的分布.

13.4 (13.2 节) 证明雅可比行列式 $|\partial(\boldsymbol{G},\boldsymbol{A})/\partial(\boldsymbol{E},\boldsymbol{F})|$ 是 $\prod(f_i-f_j)$ 乘 $\boldsymbol{E}$ 的函数, 通过证明对 $f_i=f_j$ 雅可比行列式渐变为零, f_i 中它的次数同 $\prod(f_i-f_j)$ 的相同.

13.5 (13.3 节) 用角的余弦和正弦的方法显式地给出 2×2 正交矩阵的 Haar 不变分布.

13.6 (13.3 节) 令 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是各自服从 $W(\boldsymbol{\Sigma},m)$ 和 $W(\boldsymbol{\Sigma},n)$ 分布的. 令 $l_1>\cdots>l_p$ 是 $|\boldsymbol{A}-l\boldsymbol{B}|=0$ 的根, $m_1>\cdots>m_p$ 是 $|\boldsymbol{A}-m\boldsymbol{\Sigma}|=0$ 的根. 求当 $n\to\infty$ 时从这些 l 的分布得到的 m 的分布.

13.7 (13.3 节) 用定理 13.3.1 的细节类似地证明引理 13.3.1.

13.8 令 $\boldsymbol{A}$ 是服从 $W(\boldsymbol{\Sigma},n)$ 分布的. 在 $p=2$ 的情况下求 $\boldsymbol{A}$ 的特征根的分布. [提示: 变换 $\boldsymbol{\Sigma}$ 到对角矩阵.]

13.9 从习题 13.6 的结果求球形准则的分布 (当原假设不真时).

13.10 (13.3 节) 证明 $\boldsymbol{X}$ $(p\times n)$ 有密度 $f_X(\boldsymbol{X}'\boldsymbol{X})$ 当且仅当 $\boldsymbol{T}$ 有密度

$$\frac{2^p\pi^{pn/2}}{\Gamma_p(n/2)}\prod_{i=1}^{p}t_{ii}^{n-i}f_X(\boldsymbol{T}\boldsymbol{T}'),$$

其中 $\boldsymbol{T}$ 是下三角矩阵, 有正的对角元, 使得 $\boldsymbol{T}\boldsymbol{T}'=\boldsymbol{X}'\boldsymbol{X}$. [Srivastava and Khatri(1979)]. [提示: 比较引理 13.3.1 和推论 7.2.1]

13.11 (13.5.2 节) 在协方差阵是 (12) 式的情况下求 $\boldsymbol{D}_1,\boldsymbol{W}_{11},\boldsymbol{W}_{12}$ 和 $\boldsymbol{W}_{21}$ 的极限分布.

13.12 (13.3 节) 证明 12.4 节的 (6) 式.

第14章　因子分析

14.1　引　　言

因子分析是基于把观测向量分成不可观测*系统*部分和不可观测*误差*部分的模型的. 误差向量的分量被当作不相关的或者独立的, 系统部分被当作不可观测的因子变量相对较小数目的线性组合. 因子分析把主要关心的因子效应从带有误差的数据中分离出来. 从另一个观点来看, 因子分析给出了一组变量的相依性的一个描述或者解释, 只就因子而言而不考虑观测的变异性. 而主成分分析则描述或者 "解释" 了观测的*变异性*. 因子分析起源于对智力测试的得分分析. 然而, 这个方法可以用于更广泛的领域, 比如分析意向测试组, 生理指标组, 经济数据组. 当对一群个体进行一组测试, 可以观察到, 特定个体在给定测试中的得分与他在其他测试中的得分更多地相关, 而与其他个体在其他测试中的相关较小, 即特殊个体的一些得分通常在某种程度是相互关系的. 这种相互关系可以如下解释, 即把个体的测试得分看作有两部分组成：一部分是特定测试所特有的, 称为误差; 另一部分是更多的基本量的函数, 称为*主技能得分*或者*因子得分*. 由于它们涉及好几项测试得分, 它们的效应把不同的测试得分联系起来. 大体上, 其思想就是一个在某方面更有智力的人比那些这方面差点的人将会在许多测试中都表现得更好.

因子分析的模型已在 14.2 节定义和讨论. 参数的极大似然估计量在因子得分和误差是正态分布的情况下导出, 并导出了模型拟合的检验. 对估计量和检验准则给出了大样本分布理论 (14.3 节). 固定因子的极大似然估计量是不存在的, 但是给出了相应的估计方法 (14.4 节). 解释的某些方面在 14.5 节论述. 给出了当因子是正态和识别被给定的零载荷影响时的极大似然估计量. 最后考虑了因子得分的估计. Anderson(1984a) 讨论了因子分析和主成分分析之间的关系, 以及与线性函数和结构的关系.

14.2　模　　型

14.2.1　模型的定义

令观测向量 $\boldsymbol{X}$ 为

$$\boldsymbol{X} = \boldsymbol{\Lambda f} + \boldsymbol{U} + \boldsymbol{\mu}, \tag{1}$$

其中 $\boldsymbol{X}, \boldsymbol{U}$ 和 $\boldsymbol{\mu}$ 是 p 元列向量, $\boldsymbol{f}$ 是 $m(\leqslant p)$ 元列向量, $\boldsymbol{\Lambda}$ 是 $p \times m$ 矩阵. 我们假定 $\boldsymbol{U}$ 和 $\boldsymbol{f}$ 独立地分布, 有均值 $\mathrm{E}(\boldsymbol{U}) = \boldsymbol{0}$ 和协方差阵 $\mathrm{E}(\boldsymbol{U}\boldsymbol{U}') = \boldsymbol{\Psi}$, $\boldsymbol{\Psi}$ 是对角的. 向量 $\boldsymbol{f}$ 将被可选择地视为一个随机向量或者从观测到观测变化的参数向量.

就智力测试而言, $\boldsymbol{X}$ 的每个分量是一个测试得分或者一组测试的得分. 对应的 $\boldsymbol{\mu}$ 的分量是总体中这个测试的平均得分. $\boldsymbol{f}$ 的分量是智力因子的得分, 这些分量的线性组合记入了测试得分. 这些线性组合的系数是 $\boldsymbol{\Lambda}$ 的元, 这些称作*因子载荷*. 有时 $\boldsymbol{f}$ 的元称为*公共因子*, 由于它们对一些不同的测试是公共的; 这类模型的第一类表现 [Spearman(1904)] 是 $\boldsymbol{f}$ 由一个元组成, 称这个元为*一般因子*. $\boldsymbol{U}$ 的一个元是测试得分的部分并不是由公共因子 "解释的". 这个被看作由测试的测量误差加上一个*特殊因子*组成, 特殊因子仅与这个特定测试有关. 因为在我们的模型中 (每个个体有一组观测) 我们不能区别在 $\boldsymbol{U}$ 坐标系下的两个元, 我们简称 $\boldsymbol{U}$ 的元为测量误差.

$\boldsymbol{X}$ 的给定元类似于在回归理论 (或者方差分析) 中其他变量的线性组合. 尽管在这里并没有观察到 $\boldsymbol{f}$ 扮演了独立变量的角色.

我们可以区分两类模型. 在一个中我们认为向量 $\boldsymbol{f}$ 是随机向量, 在另一个中我们认为 $\boldsymbol{f}$ 是从一个个体到另一个个体变化的非随机量的向量. 在第二种情况下, 写成 $\boldsymbol{X}_\alpha = \boldsymbol{\Lambda}\boldsymbol{f}_\alpha + \boldsymbol{U} + \boldsymbol{\mu}$ 更准确. 非随机因子得分向量似乎是一个系统部分更好的描述, 但是推理变得困难, 因为似然函数可能没有极大值. 原则上, 当不同的样本包括不同的个体时包括随机因子的模型更恰当; 当包括特殊个体并且不只对结构感兴趣时非随机因子模型更合适.

当把 $\boldsymbol{f}$ 视为随机的时, 我们假定 $\mathrm{E}(\boldsymbol{f}) = \boldsymbol{0}$. (否则, $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\Lambda}\mathrm{E}(\boldsymbol{f}) + \boldsymbol{\mu}$, 可以重新定义 $\boldsymbol{\mu}$, 使其合并 $\boldsymbol{\Lambda}\mathrm{E}(\boldsymbol{f})$.) 令 $\mathrm{E}(\boldsymbol{f}\boldsymbol{f}') = \boldsymbol{\Phi}$. 我们的分析将用一阶矩和二阶矩的方法. 通常, 我们认为 $\boldsymbol{f}$ 和 $\boldsymbol{U}$ 有正态分布. 如果 $\boldsymbol{f}$ 不是随机的, 则对第 α 个个体有 $\boldsymbol{f} = \boldsymbol{f}_\alpha$. 于是我们可以假定通常有 $(1/N)\sum_{\alpha=1}^{N} \boldsymbol{f}_\alpha = \boldsymbol{0}$ 和 $(1/N)\sum_{\alpha=1}^{N} \boldsymbol{f}_\alpha \boldsymbol{f}'_\alpha = \boldsymbol{\Phi}$.

这个模型中有函数的不确定性. 令 $\boldsymbol{f} = \boldsymbol{C}\boldsymbol{f}^*(\boldsymbol{f}^* = \boldsymbol{C}^{-1}\boldsymbol{f})$ 和 $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{C}$, 其中 $\boldsymbol{C}$ 是非奇异 $m \times m$ 矩阵. 则 (1) 可以写成

$$\boldsymbol{X} = \boldsymbol{\Lambda}^*\boldsymbol{f}^* + \boldsymbol{U} + \boldsymbol{\mu}. \tag{2}$$

当 $\boldsymbol{f}$ 是随机的时, $\mathrm{E}(\boldsymbol{f}^*\boldsymbol{f}^{*'}) = \boldsymbol{C}^{-1}\boldsymbol{\Phi}(\boldsymbol{C}^{-1})' = \boldsymbol{\Phi}^*$; 当 $\boldsymbol{f}$ 是非随机的时, $(1/N)\sum_{\alpha=1}^{N} \boldsymbol{f}^*_\alpha \boldsymbol{f}^{*'}_\alpha = \boldsymbol{\Phi}^*$. 带有 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{f}$ 的模型等价于带有 $\boldsymbol{\Lambda}^*$ 和 $\boldsymbol{f}^*$ 的模型, 即我们不能通过观察 $\boldsymbol{X}$ 来区分这两个模型.

通过规定 $\mathrm{E}(\boldsymbol{f}\boldsymbol{f}') = \boldsymbol{I}$(如果 $\boldsymbol{f}$ 是随机的) 或者 $\sum_{\alpha=1}^{N} \boldsymbol{f}_\alpha \boldsymbol{f}'_\alpha = N\boldsymbol{I}$(如果 $\boldsymbol{f}$ 不是随机的) 可以消除模型中的一些不确定性. 这种情况下称因子为*正交的*; 如果 $\boldsymbol{\Phi}$ 不是对角的, 称因子是*偏斜的*. 当我们假定 $\boldsymbol{\Phi} = \boldsymbol{I}$ 时, 则有 $\mathrm{E}(\boldsymbol{f}^*\boldsymbol{f}^{*'}) = \boldsymbol{C}^{-1}(\boldsymbol{C}^{-1})' = \boldsymbol{I}(\boldsymbol{I} = \boldsymbol{C}\boldsymbol{C}')$. 不确定性等价于乘一个正交矩阵, 这称为*旋转*问题. 要求 $\boldsymbol{\Phi}$ 是对角的

意味着当假定 f 是正态的时 f 的元是独立分布的. 这个引起心理学家的兴趣, 因为一个公共的智力因子的思想就是 (由定义) 它们是独立的或者是不相关的量.

一个极其重要的假设是 U 的元是不相关的. 我们的观点是观测误差和特殊因子按照定义是不相关的. 即测试得分的相互关系是公共因子引起的, 那正是我们想要研究的. 关于因子分析还有另一种从根本上就完全不同的观点, 即假设公共因子尽可能多地解释或者说明测试得分的方差. 为了理解这种观点, 我们应该用一个不同的模型.

几何描述帮助我们建立直觉. 考虑一个 p 维空间. $\boldsymbol{\Lambda}$ 的列可以被当作这个空间的 m 维向量. 它们张成了一些 m 维的子空间; 事实上, 它们可以被当作 m 维空间的坐标系, f 可以被当作在那个空间中提到的这个特殊的轴系统下一个点的坐标. 称这个子空间是**因子空间**. $\boldsymbol{\Lambda}$ 右端乘一个在因子空间中相应的取一组新的坐标轴的矩阵.

如果因子是随机的, 观测 $\boldsymbol{X}$ 的协方差阵是

$$\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})' = \mathrm{E}(\boldsymbol{\Lambda f}+\boldsymbol{U})(\boldsymbol{\Lambda f}+\boldsymbol{U})' = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}. \tag{3}$$

如果因子是正交的 ($\mathrm{E}(ff') = \boldsymbol{I}$), 则 (3) 是

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Lambda}' + \boldsymbol{\Psi}. \tag{4}$$

如果 f 和 $\boldsymbol{U}$ 是正态的, 所有关于结构的信息都来自于 (3)[或 (4)] 还有 $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}$.

14.2.2 识别

给定协方差阵 $\boldsymbol{\Sigma}$ 和 m 个因子, 我们可以看看是否存在一个三元组 $\boldsymbol{\Lambda}, \boldsymbol{\Phi}$(正定), $\boldsymbol{\Psi}$(正定对角) 满足 (3) 式. 如果存在, 那三元组是唯一的吗? 因为任意三元组可以变换到一个等价结构 $\boldsymbol{\Lambda C}, \boldsymbol{C}^{-1}\boldsymbol{\Phi C}'^{-1}, \boldsymbol{\Psi}$, 我们可以把 m^2 个独立条件加在 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Phi}$ 上来排除这个不确定性. 观测 $\boldsymbol{\Sigma}$ 的元的个数和条件的个数 (唯一性) 是 $\frac{1}{2}p(p+1)+m^2$; $\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$ 中的参数个数分别是 $pm, \frac{1}{2}m(m+1), p$. 如果观测数量和条件对参数个数的超出量, 也就是, $\frac{1}{2}[(p-m)^2-p-m]$, 是正的, 我们可以排除存在性但是如果一组参数确实存在则有唯一性. 如果超出量是负的, 我们可以认为存在但是有可能不唯一; 如果超过量是 0, 我们希望存在并且唯一 (或者至少有有限组解). 解的存在性问题也就是是否存在一个对角矩阵 $\boldsymbol{\Psi}$, 有非负的对角元使得 $\boldsymbol{\Sigma}-\boldsymbol{\Psi}$ 是半正定的, 且秩为 m. Anderson and Rubin(1956) 包括关于这个问题的大部分已知的结果.

如果解存在并且唯一, 模型称为**可识别的**. 正如上面提到的, 必须加 m^2 个条件于 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Phi}$ 上来消除变换 $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda C}$ 和 $\boldsymbol{\Phi}^* = \boldsymbol{C}^{-1}\boldsymbol{\Phi C}'^{-1}$. 我们上面提到条件 $\boldsymbol{\Phi} = \boldsymbol{I}$, 它使得变换 $\boldsymbol{C}$ 是正交的. [在 $\boldsymbol{\Phi} = \boldsymbol{I}$ 中有 $\frac{1}{2}m(m+1)$ 个分量方程.]为了某些目的, 加上下面的约束条件更方便,

$$\boldsymbol{\Gamma} = \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} \tag{5}$$

是对角的. 如果 $\mathbf{\Gamma}$ 的对角元是有序的并且不相同的 $(\gamma_{11} > \gamma_{22} > \cdots > \gamma_{mm})$, 则 $\mathbf{\Lambda}$ 是唯一确定的. 另一个条件是 $\mathbf{\Lambda}$ 的前 m 行形成下三角矩阵. 这个条件的一般化是要求 $\boldsymbol{B\Lambda}$ 的前 m 行形成下三角矩阵, 其中 $\boldsymbol{B}$ 是前面给出的. (这个条件被所谓的形心方法包含.)

简单结构

Thurstone(1947, p. 335) 提出了一些条件, 在类 $\mathbf{\Lambda}\boldsymbol{C}$ 外选择一个矩阵使得其有特殊的心理学的意义. 如果 $\lambda_{i\alpha} = 0$, 则第 α 个因子没有记入第 i 个测试. **简单结构**的一般思想是当因子有实际的心理学意义时, 很多测试不取决于所有的因子. 这表明, 给出 $\mathbf{\Lambda}$, 需要考虑所有的旋转, 即其中 $\boldsymbol{C}$ 是正交的所有的矩阵 $\mathbf{\Lambda}\boldsymbol{C}$, 从中选择出有最多 0 系数的. 可以认为这个矩阵给出最简单的结构, 假定它有最有意义的心理学解释. 心理学家可以构造他的测试, 使得它们以不同的方式取决于假定的因子.

这些 0 的位置并没有预先选择, 但是尝试旋转 $\boldsymbol{C}$ 直到找到满足这些条件的 $\mathbf{\Lambda}$. 这些条件效应的识别并不清楚. Reiersøl(1950) 修改了 Thurstone 的条件, 使得只有一种旋转满足条件, 因此影响识别.

指定位置的零元素

这里我们考虑需要调查者提供更多先验信息的一组条件. 调查者必须知道一些特殊测试不依靠于某些特殊因子. 在这种情况下, 对于特殊的组 (i, α), 条件是 $\lambda_{i\alpha} = 0$, 即第 α 个因子不影响第 i 个测试得分. 则我们不假设 $\mathrm{E}(\boldsymbol{ff}') = \boldsymbol{I}$. 这些条件类似于经济模型中用的一些条件. 第 α 列的系数被识别, 除了乘一个比例因子, 如果 (a) 那个列有至少 $m-1$ 个零元素和如果 (b)$\mathbf{\Lambda}^{(\alpha)}$ 的秩是 $m-1$, 其中 $\mathbf{\Lambda}^{(\alpha)}$ 是行包括指定的 0 且第 α 列删除指定的 0 的矩阵 (比如, 第 α 列删除). (见习题 14.1) 列乘一个比例常数可以由正规化消除, 比如对每个 α 和某些 i, $\phi_{\alpha\alpha} = 1$ 或者 $\lambda_{i\alpha} = 1$. 如果 $\phi_{\alpha\alpha} = 1, \alpha = 1, \cdots, m$, 则 $\mathbf{\Phi}$ 是一个相关阵.

将会看出有 m 个正规化和最小有 $m(m-1)$ 零条件. 这和 $\boldsymbol{C}$ 的元的数量是相等的. 如果在指定的 $\mathbf{\Lambda}$ 的一列或者多列有多于 $m-1$ 个零元, 则消除 $\mathbf{\Lambda}\boldsymbol{C}$ 的不确定性可能需要更多的条件; 在这种情况下条件可能限制 $\mathbf{\Lambda\Phi\Lambda}'$.

作为一个例子, 对于五个测试的得分, 考虑模型

$$\boldsymbol{X} = \boldsymbol{\mu} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & \lambda_{32} \\ 0 & \lambda_{42} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ a \end{bmatrix} + \boldsymbol{U}$$

$$= \boldsymbol{\mu} + \begin{bmatrix} v \\ \lambda_{21}v \\ \lambda_{31}v + \lambda_{32}a \\ \lambda_{42}a \\ a \end{bmatrix} + \boldsymbol{U} \tag{6}$$

其中 v 和 a 是语言和算术能力的度量. 当最后两组测试仅仅依赖于算术能力时, 指定前两组试验仅仅依赖于语言能力. 正规化将语言能力用第一个测试度量将算术能力用第五个测试度量.

Koopmans and Reiersøl(1950), Anderson and Rubin(1956), 还有 Howe(1955) 给出先前指定的 0 在识别中的应用和这个情况在正态下的极大似然估计. [见 Lawley(1958).] Jöreskog(1969) 称因子分析在这些识别条件下为**确证性**因子分析; 带有任意的条件或者到简单的结构的旋转的因子分析, 称为**探索性因子分析**.

其他条件

一组方便的条件是要求 $\boldsymbol{\Lambda}$ 的上子方矩阵是已给定的. 这假设了没有这个条件的上方阵是非奇异的. 事实上, 如果 $\boldsymbol{\Lambda}^* = (\boldsymbol{\Lambda}_1^{*'}, \boldsymbol{\Lambda}_2^{*'})'$ 是一个任意 $p \times m$ 矩阵, $\boldsymbol{\Lambda}_1^*$ 是方阵且非奇异, 则 $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}_1^{*-1} = (\boldsymbol{I}_m, \boldsymbol{\Lambda}_2')'$ 满足条件. (把 $\boldsymbol{\Lambda}$ 的 $m \times m$ 主子阵作为 $\boldsymbol{I}_m$ 只是一个方便的识别条件, 并没有给出任何实质性意义.)

14.2.3 度量的单位

我们考虑把因子分析方法用于协方差阵. 许多情况下 $\boldsymbol{X}$ 的每个元的度量单位是任意的. 例如, 生理测试中得分的单位没有实质的意义.

变换度量单位意味着 $\boldsymbol{X}$ 的每个元乘一个常数; 这些常数不必相等. 当一个给定的测试得分乘一个常数时, 测试的因子载荷乘相同的常数, 误差方差乘常数的平方. 假设 $\boldsymbol{DX} = \boldsymbol{X}^*$, 其中 $\boldsymbol{D}$ 是对角阵且有正的对角元. 则 (1) 变成

$$\boldsymbol{X}^* = \boldsymbol{\Lambda}^*\boldsymbol{f} + \boldsymbol{U}^* + \boldsymbol{\mu}^*, \tag{7}$$

其中 $\boldsymbol{\mu}^* = \mathrm{E}(\boldsymbol{X}^*) = D\boldsymbol{\mu}, \boldsymbol{\Lambda}^* = \boldsymbol{D\Lambda}, \boldsymbol{U}^* = \boldsymbol{DU}$ 有协方差阵 $\boldsymbol{\Psi}^* = \boldsymbol{D\Psi D}$. 则

$$\mathrm{E}(\boldsymbol{X}^* - \boldsymbol{\mu}^*)(\boldsymbol{X}^* - \boldsymbol{\mu}^*)' = \boldsymbol{\Lambda}^*\boldsymbol{\Phi}\boldsymbol{\Lambda}^{*'} + \boldsymbol{\Psi}^* = \boldsymbol{\Sigma}^*, \tag{8}$$

其中 $\boldsymbol{\Sigma}^* = \boldsymbol{D\Sigma D}$. 注意到, 如果识别条件是 $\boldsymbol{\Phi} = \boldsymbol{I}$ 和 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 为对角的, 则 $\boldsymbol{\Lambda}^*$ 满足后一条件. 如果 $\boldsymbol{\Lambda}$ 是由给定的 0 识别并且通过 $\Phi_{\alpha\alpha} = 1 (\alpha = 1, \cdots, m)$ (例如, $\boldsymbol{\Phi}$ 是相关阵) 正规化, 则 $\boldsymbol{\Lambda}^* = \boldsymbol{D\Lambda}$ 可类似地识别. (如果对指定的 i 和每个 α 正规化是 $\lambda_{i\alpha} = 1$, 则 $\boldsymbol{D\Lambda}$ 的每列必须再正规化.)

一个特殊的对角矩阵 $\boldsymbol{D}$ 包括观测标准差的倒数 $d_{ii} = 1/\sqrt{\sigma}_{ii}$. 则 $\boldsymbol{\Sigma}^* = \boldsymbol{D\Sigma}\mathrm{D}$ 是相关阵.

我们一会将要看到关于识别条件 $\boldsymbol{\Gamma}$ 为对角或者指定的 0 变换的极大似然估计量, 即是变换 $\boldsymbol{x}_\alpha^* = \boldsymbol{D}\boldsymbol{x}_\alpha, \alpha = 1, \cdots, N$, 使得 $\hat{\boldsymbol{\Lambda}}^* = \boldsymbol{D}\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}^* = \boldsymbol{D}\hat{\boldsymbol{\Psi}}\boldsymbol{D}$.

14.3 随机正交因子的极大似然估计量

14.3.1 极大似然估计量

本节我们求得当观测是正态分布时的参数的极大似然估计量, 即因子得分和误差是正态的 [Lawley(1940)]. 则 $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$. 我们强加条件于 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Phi}$ 上使得它们被唯一识别. 这些没有限制 $\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'$, 它是正定矩阵且有秩 m. 为了方便我们假设 $\boldsymbol{\Phi} = \boldsymbol{I}$(比如因子是正交的或者不相关的) 和 $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的. 则似然取决于均值 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$. 在其他一些影响识别的条件下, $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Phi}$ 的极大似然估计量 [比如, $\boldsymbol{\Lambda} = (\boldsymbol{I}_m, \boldsymbol{\Lambda}_2')'$] 是 $\boldsymbol{\Lambda}$ 在先前的条件下的极大似然估计量的变换. 如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是对 $\boldsymbol{X}$ 的 N 个的观测, 样本的似然函数是

$$L = (2\pi)^{-\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{-\frac{1}{2}N} \exp\left[-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right]. \tag{1}$$

均值 $\boldsymbol{\mu}$ 的极大似然估计量是 $\hat{\boldsymbol{\mu}} = \bar{x} = (1/N)\sum_{\alpha=1}^{N}\boldsymbol{x}_\alpha$.

令

$$\boldsymbol{A} = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})'. \tag{2}$$

下一步我们极大化 (1) 式的对数用 $\hat{\boldsymbol{\mu}}$ 取代 $\boldsymbol{\mu}$, 即 [①]

$$-\frac{1}{2}pN\ln 2\pi - \frac{1}{2}N\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{tr}\,\boldsymbol{A}\boldsymbol{\Sigma}^{-1}. \tag{3}$$

(这是集中似然的对数.) 从 $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} = \boldsymbol{I}$, 我们得到, 对任意参数 θ,

$$\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\theta} = -\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta}\boldsymbol{\Sigma}^{-1}. \tag{4}$$

则 (3) 式关于 ψ_{ii} 的偏导数, 即 $\boldsymbol{\Psi}$ 的对角元, 是 $-N/2$ 倍的

$$\sigma^{ii} - \sum_{k,j=0}^{p} c_{kj}\sigma^{ji}\sigma^{ik}, \tag{5}$$

其中 $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$ 和 $(c_{ij}) = \boldsymbol{C} = (1/N)\boldsymbol{A}$. 用矩阵的记法, 令 (5) 式等于 0 得到

$$\mathrm{diag}\boldsymbol{\Sigma}^{-1} = \mathrm{diag}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}^{-1}, \tag{6}$$

其中 $\mathrm{diag}\boldsymbol{H}$ 指定了矩阵 $\boldsymbol{H}$ 的对角项. 等价地, $\mathrm{diag}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} - \boldsymbol{C})\boldsymbol{\Sigma}^{-1} = \mathrm{diag}\boldsymbol{0}$. (3) 式关于 $\lambda_{k\tau}$ 的导数是 $-N$ 倍

① 我们应该加上约束使得 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 用拉格朗日乘子的非对角元是 0, 但是令导数等于 0 时拉格朗日乘子变为 0. 这些约束并不影响极大值.

$$\sum_{j=1}^{p} \sigma^{kj}\lambda_{j\tau} - \sum_{h,g,j=1}^{p} \sigma^{kh}c_{hg}\sigma^{gi}\lambda_{j\tau},\ k=1,\cdots,p,\ \tau=1,\cdots,m. \tag{7}$$

用矩阵的记法, 令 (7) 式等于 0 得出

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}. \tag{8}$$

我们有

$$\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{\Gamma} + \boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\Gamma} + \boldsymbol{I}). \tag{9}$$

由此我们得到 $\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}(\boldsymbol{\Gamma}+\boldsymbol{I})^{-1} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}$. 从而用 $\boldsymbol{\Sigma}$ 乘 (8) 式可以得到

$$\boldsymbol{\Lambda}(\boldsymbol{\Gamma}+\boldsymbol{I}) = \boldsymbol{C}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}, \tag{10}$$

或者

$$(\boldsymbol{C} - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{\Gamma}. \tag{11}$$

下面我们想要证明, 当 (8) 式成立时 $\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Sigma}^{-1}$ 就是 $\boldsymbol{\Psi}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Psi}^{-1}$. 将后者左端和右端同时乘 $\boldsymbol{\Sigma}$ 得到

$$\begin{aligned}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma} &= (\boldsymbol{\Lambda}\boldsymbol{\Lambda}'+\boldsymbol{\Psi})\boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'-\boldsymbol{C})\boldsymbol{\Psi}^{-1}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'+\boldsymbol{\Psi}) \\ &= \boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}' - \boldsymbol{C},\end{aligned} \tag{12}$$

这是因为由 (10) 式有

$$\begin{aligned}\boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'-\boldsymbol{C}) &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Lambda}' - \boldsymbol{\Lambda}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{C} \\ &= \boldsymbol{\Lambda}[(\boldsymbol{I}+\boldsymbol{\Gamma})\boldsymbol{\Lambda}' - \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{C}] \\ &= \boldsymbol{0}.\end{aligned} \tag{13}$$

因此

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Psi}^{-1}. \tag{14}$$

则 (6) 式等价于 $\operatorname{diag}\boldsymbol{\Psi}^{-1}(\boldsymbol{\Sigma}-\boldsymbol{C})\boldsymbol{\Psi}^{-1} = \operatorname{diag}\boldsymbol{0}$. 因为 $\boldsymbol{\Psi}$ 是对角的, 这个方程等价于

$$\operatorname{diag}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}'+\boldsymbol{\Psi}) = \operatorname{diag}\boldsymbol{C}. \tag{15}$$

$\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}$ 的估计量是由 (10) 式和 (15) 式确定的, 要求 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的.

我们可以将 (11) 式左端乘 $\boldsymbol{\Psi}^{-\frac{1}{2}}$, 得到

$$\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{C}-\boldsymbol{\Psi})\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Lambda}) = (\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Lambda})\boldsymbol{\Gamma}, \tag{16}$$

这表明 $\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Lambda}$ 的列是 $\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{C}-\boldsymbol{\Psi})\boldsymbol{\Psi}^{-\frac{1}{2}} = \boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}} - \boldsymbol{I}$ 的特征向量, $\boldsymbol{\Gamma}$ 相应的对角元是特征根. [事实上, $\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}} - \boldsymbol{I}$ 的特征向量是 $\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}}$ 的特征向量, 因为 $(\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}} - \boldsymbol{I})\boldsymbol{x} = \gamma\boldsymbol{x}$ 是等价于 $\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{x} = (1+\gamma)\boldsymbol{x}$.] 向量由 $(\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Lambda})'(\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{\Lambda}) = \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda} = \boldsymbol{\Gamma}$ 正规化. 选择特征根来极大化似然. 为了估计极大化似然函数, 我们计算

$$\begin{aligned}\operatorname{tr} \boldsymbol{C}\hat{\boldsymbol{\Sigma}}^{-1} &= \operatorname{tr} \boldsymbol{C}\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}')\hat{\boldsymbol{\Psi}}^{-1} \\ &= \operatorname{tr}\left[\boldsymbol{C}\hat{\boldsymbol{\Psi}}^{-1} - (\boldsymbol{C}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Lambda}})\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-1}\right] \\ &= \operatorname{tr}\left[\boldsymbol{C}\hat{\boldsymbol{\Psi}}^{-1} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-1}\right] \\ &= \operatorname{tr}\left[(\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-1} - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-1}\right] \\ &= p.\end{aligned} \tag{17}$$

第三个等式是由 (8) 式左端乘 $\hat{\boldsymbol{\Sigma}}$ 得到的; 第四个等式是由 (15) 式和 $\hat{\boldsymbol{\Psi}}$ 是对角的事实得到的. 下面我们求

$$\begin{aligned}|\hat{\boldsymbol{\Sigma}}| &= |\hat{\boldsymbol{\Psi}}^{\frac{1}{2}}| \cdot |\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}} + \boldsymbol{I}_p| \cdot |\hat{\boldsymbol{\Psi}}^{\frac{1}{2}}| \\ &= |\hat{\boldsymbol{\Psi}}| \cdot |\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\hat{\boldsymbol{\Lambda}} + \boldsymbol{I}_m| \\ &= |\hat{\boldsymbol{\Psi}}| \cdot |\hat{\boldsymbol{\Gamma}} + \boldsymbol{I}_m| \\ &= \prod_{i=1}^{p} \hat{\psi}_{ii} \prod_{j=1}^{m} (\hat{\gamma}_j + 1).\end{aligned} \tag{18}$$

对于 $\boldsymbol{U}_{p\times m}$, 第二个等式是 $|\boldsymbol{U}\boldsymbol{U}' + \boldsymbol{I}_p| = |\boldsymbol{U}'\boldsymbol{U} + \boldsymbol{I}_m|$, 这已经在 8.4 节的 (14) 式证明过. 从事实 $\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{C} - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-\frac{1}{2}}$ 的特征根是 $0 = |\boldsymbol{C} - \boldsymbol{\Psi} - \gamma\boldsymbol{\Psi}| = |\boldsymbol{C} - (1+\gamma)\boldsymbol{\Psi}|$ 的根 $\gamma_1 > \gamma_2 > \cdots > \gamma_p$, 可得

$$\frac{|\boldsymbol{C}|}{|\hat{\boldsymbol{\Psi}}|} = \prod_{i=1}^{p}(1 + \hat{\gamma}_i). \tag{19}$$

[注意到 $\boldsymbol{\Psi}^{-\frac{1}{2}}\boldsymbol{C}\boldsymbol{\Psi}^{-\frac{1}{2}}$ 的根 $1+\gamma_i$ 是正的. $\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{C} - \boldsymbol{\Psi})\boldsymbol{\Psi}^{-\frac{1}{2}}$ 的根 γ_i 不必是正的, 通常有一些会是负的.]

于是

$$|\hat{\boldsymbol{\Sigma}}| = \frac{|\boldsymbol{C}| \prod_{j\in S}(1 + \hat{\gamma}_j)}{\prod_{i=1}^{p}(1 + \hat{\gamma}_i)} = \frac{|\boldsymbol{C}|}{\prod_{j\notin S}(1 + \hat{\gamma}_j)}, \tag{20}$$

其中 S 是与 $\hat{\boldsymbol{\Gamma}}$ 中的根相对应的一组指标. 极大化似然函数的对数是

$$-\frac{1}{2}pN\ln 2\pi - \frac{1}{2}N\ln|\boldsymbol{C}| - \frac{1}{2}N\sum_{j\notin S}\ln(1 + \hat{\gamma}_j) - \frac{1}{2}Np. \tag{21}$$

最大的根 $\hat{\gamma}_1 > \cdots > \hat{\gamma}_m$ 可以被选择作为 $\hat{\boldsymbol{\Gamma}}$ 的对角元. 则 $S = \{1, \cdots, m\}$. 集中似然 (3) 式的对数是 $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ 的函数. 对于每个 $\boldsymbol{\Lambda}$ 和每个对角的正定矩阵 $\boldsymbol{\Psi}$, 这个矩阵是正定的, 对于一些对角的非正定矩阵 $\boldsymbol{\Psi}$, 它同样是正定的. 因此对正定的 $\boldsymbol{\Psi}$ 未必相对极大值. 集中似然函数可能随着 $\boldsymbol{\Psi}$ 的一个或者多个对角元趋近 0 而增大. 在那种情况下, 导数方程对正定的 $\boldsymbol{\Psi}$ 可能不满足.

估计量 (11) 和 (15) 的方程可以写成多项式方程 [$|\boldsymbol{\Psi}|$ 乘 (11) 式], 但是不能直接解出. 求得似然函数的极大值有各种迭代的方法, 包括最速下降, 牛顿–拉弗森, 得分 (利用信息阵) 和 Fletcher-Powell. [见 Lawley and Maxwell(1971), Appendix II.]

因为在区域 $\psi_{ii} > 0(i = 1, \cdots, p)$ 中可能没有一个相对的极大值, 一个迭代步骤可能定义 $\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}$ 的一系列值, 包括对某些指标 i, $\hat{\psi}_{ii} < 0$. 这样的负值是不容许的, 因为 ψ_{ii} 被解释为误差的方差. 可能强加条件 $\psi_{ii} \geqslant 0, i = 1, \cdots, p$. 则极大值可能在边缘出现 (不一定满足所有的导数方程). 对某些指标 i, 误差的估计方差是 0, 即一些测试得分恰好是因子得分的线性组合. 如果忽略识别条件 $\boldsymbol{\Phi} = \boldsymbol{I}$ 和 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的, 我们可以找到一组坐标系使得有 0 误差方差的测试得分因子被解释为 (变换的) 因子得分. 这个解释看起来并没有多大用处.[进一步的讨论见 Lawley and Maxwell (1971).]

要求 ψ_{ii} 是正的另一种方式是要求 ψ_{ii} 离 0 有距离. 一个可能是对某些较小的 ε, 比如 0.005, $\psi_{ii} \geqslant \varepsilon\sigma_{ii}$. 当然, ε 的值是任意的; 如果极大值不在限制区域的内部, ε 的增大会使极大值减小, 导数方程将不会全部满足.

集中似然的本质是可能有多于一个的相对极大值. 其中极大值由迭代方法近似, 它取决于初始值.Rubin and Thayer (1982) 用 EM 算法给出从三组不同的初始值出发的三组估计.

EM(期望极大化) 算法是对极大似然估计的一个可能的算法 [Dempster, Laird, and Rubin (1977) 和 Rubin and Thayer (1982)]. 算法的思想是将没有观测的 $\boldsymbol{f}$ 视为缺失数据. 在假设 f 和 U 有联合正态分布时, 充分统计量是这些 $\boldsymbol{X}$ 和 $\boldsymbol{f}$ 的均值和协方差. 算法的 E 步是基于参数的试验值得到协方差的期望值. M 步是基于这些协方差极大化似然函数; 这个步骤提供了更新的参数值. 步骤交替, 该方法通常收敛到极大似然估计量.(见习题 14.3 .)

正如 14.2 节中提到的, 结构是等价的, 因子得分在观测变量的度量单位的变换 $\boldsymbol{X} \to \boldsymbol{DX}$ 下是不变量, 其中 $\boldsymbol{D}$ 是对角矩阵, 有正的对角元, $\boldsymbol{\Lambda}$ 由 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的识别. 如果我们令 $\boldsymbol{D\Lambda} = \boldsymbol{\Lambda}^*, \boldsymbol{D\Psi D} = \boldsymbol{\Psi}^*$, $\boldsymbol{DCD} = \boldsymbol{C}^*$, 则似然函数的对数是一个常数乘一个常数倍的

$$\begin{aligned} &-\ln|\boldsymbol{\Psi}^* + \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*'}| - \operatorname{tr} C^*(\boldsymbol{\Psi}^* + \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*'})^{-1} \\ &= -\ln|\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}'| - \operatorname{tr} C(\boldsymbol{\Psi} + \boldsymbol{\Lambda}\boldsymbol{\Lambda}')^{-1} - 2\ln|\boldsymbol{D}|. \end{aligned} \tag{22}$$

$\boldsymbol{\Lambda}^*$ 和 $\boldsymbol{\Psi}^*$ 的极大似然估计量是 $\hat{\boldsymbol{\Lambda}}^* = \boldsymbol{D}\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}^* = \boldsymbol{D}\hat{\boldsymbol{\Psi}}\boldsymbol{D}$, $\hat{\boldsymbol{\Lambda}}^{*'}\hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{\Lambda}}^* = \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}$ 是对角的. 即估计的因子载荷和误差方差只是被度量单位改变.

常常为了方便, 利用 $d_{ii} = 1/\sqrt{c_{ii}}$, 所以 $\boldsymbol{DCD} = (r_{ij})$ 是由样本相关系数组成的. 分析和度量单位是独立的. 这个事实同心理学测试得分没有本质单位的事实是相关的.

因子不取决于位置和比例因子的事实是把因子分析视为互相依赖性分析的一个原因.

就观测的相关而言, 更方便的是给出初始估计共同的一些经验法则, $\sum_{j=1}^{m} \hat{\lambda}_{ij}^2 = 1 - \hat{\psi}_{ii}$. 一个法则是用 $R^2_{i\cdot 1,\cdots,i-1,i+1,\cdots,p}$. 另一个是用 $\max_{h \neq i}|r_{ih}|$.

14.3.2 模型拟合的假设检验

我们将讨论模型拟合的似然比检验，即对指定的 m，对某些正定的对角矩阵 $\boldsymbol{\Psi}$ 和某些 $p\times m$ 矩阵 $\boldsymbol{\Lambda}$，协方差阵可以写成 $\boldsymbol{\Sigma}=\boldsymbol{\Psi}+\boldsymbol{\Lambda\Lambda}'$. 似然比准则是

$$\frac{\max_{\mu,\Lambda,\Psi}L(\boldsymbol{\mu},\boldsymbol{\Psi}+\boldsymbol{\Lambda\Lambda}')}{\max_{\mu,\Sigma}L(\boldsymbol{\mu},\boldsymbol{\Sigma})}=\frac{|\boldsymbol{C}|^{\frac{1}{2}N}}{|\hat{\boldsymbol{\Psi}}+\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'|^{\frac{1}{2}N}}=\prod_{j=m+1}^{p}(1+\hat{\gamma}_j)^{\frac{1}{2}N}, \tag{23}$$

因为 $\boldsymbol{\Sigma}$ 的无限制极大似然估计量是 $\boldsymbol{C}$，由 (17) 式，$\operatorname{tr}\boldsymbol{C}(\hat{\boldsymbol{\Psi}}+\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}})^{-1}=p$，从 (20) 式有 $|\boldsymbol{C}|/|\hat{\boldsymbol{\Sigma}}|=\prod_{j=m+1}^{k}(1+\hat{\gamma}_j)^{\frac{1}{2}N}$. 如果 (23) 式很小则原假设被拒绝. 我们可以用 -2 倍的似然比准则的对数，即

$$-N\sum_{j=m+1}^{p}\ln(1+\hat{\gamma}_j), \tag{24}$$

如果 (24) 式太大的话则拒绝原假设.

如果 $\hat{\boldsymbol{\Psi}}$ 和 $\hat{\boldsymbol{\Lambda}}$ 的正则条件是渐近正态分布的成立. 在原假设下 (24) 式的极限分布是自由度为 $\frac{1}{2}[(p-m)^2-p-m]\chi^2$，其中自由度等于 $\boldsymbol{\Sigma}$ 的元的个数加上识别限制的个数再减去 $\boldsymbol{\Psi}$ 和 $\boldsymbol{\Lambda}$ 中的参数个数. Bartlett(1950) 指出用 $N-(2p+11)/6-2m/3$[①]代替 N. 见 Amemiya and Anderson(1990).

从 (15) 式以及 $\hat{\gamma}_1,\cdots,\hat{\gamma}_p$ 是 $\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}$ 的特征根，我们得到

$$\begin{aligned}0&=\operatorname{tr}\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}}-\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}')\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\\&=\operatorname{tr}\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}-\operatorname{tr}\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}\\&=\operatorname{tr}\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}-\operatorname{tr}\hat{\boldsymbol{\Gamma}}\\&=\sum_{i=1}^{p}\hat{\gamma}_i-\sum_{i=1}^{m}\hat{\gamma}_i=\sum_{i=m+1}^{p}\hat{\gamma}_i.\end{aligned} \tag{25}$$

如果对于 $j=m+1,\cdots,p$，$|\hat{\gamma}_j|<1$，我们可以用 (25) 式将 (24) 式展开为

$$-N\sum_{j=m+1}^{p}\left(\hat{\gamma}_j-\frac{1}{2}\hat{\gamma}_j^2+\frac{1}{3}\hat{\gamma}_j^3-\cdots\right)=\frac{1}{2}N\sum_{j=m+1}^{p}\left(\hat{\gamma}_j^2-\frac{2}{3}\hat{\gamma}_j^3+\cdots\right). \tag{26}$$

准则近似是 $\frac{1}{2}N\sum_{j=m+1}^{p}\hat{\gamma}_j^2$. 估计量 $\hat{\boldsymbol{\Psi}}$ 和 $\hat{\boldsymbol{\Lambda}}$ 使得 $\boldsymbol{C}-\hat{\boldsymbol{\Psi}}-\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}'$ 的统计灵敏性较小，或者等价地，$\boldsymbol{C}-\hat{\boldsymbol{\Psi}}$ 近似有秩 m. 则 $\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}$ 的较小的 $p-m$ 个根可能在 0 附近. 准则度量了这些根到 0 的偏差. 因为 $\hat{\gamma}_{m+1},\cdots,\hat{\gamma}_p$ 是 $\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Psi}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}$ 的非零根，我们看出

---

① 这个因子是启发式的. 如果 $m=0$，第 9 章的因子是 $N-(2p+11)/6$; Bartlett 指出分别用 $N-m$ 和 $p-m$ 取代 N 和 p.

$$\frac{1}{2}\sum_{j=m+1}^{p}\hat{\gamma}_j^2=\frac{1}{2}\mathrm{tr}\,[\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}(\boldsymbol{C}-\hat{\boldsymbol{\Sigma}})\hat{\boldsymbol{\Psi}}^{-\frac{1}{2}}]^2 \tag{27}$$

$$=\frac{1}{2}\mathrm{tr}\,\hat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{C}-\hat{\boldsymbol{\Sigma}})\hat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{C}-\hat{\boldsymbol{\Sigma}})$$

$$=\sum_{i<j}\frac{(c_{ij}-\hat{\sigma}_{ij})^2}{\hat{\psi}_{ii}\hat{\psi}_{jj}},$$

因为 $\boldsymbol{C}-\hat{\boldsymbol{\Sigma}}$ 的对角元是 0.

许多情形下研究者并不知道假设的 m 的值. 他想要确定使得模型同数据是一致的因子的最小数目. 习惯上检验 m 的连续的值. 研究者从检验一个因子的数目是一个指定的 m_0 开始 (可能是 0 或者 1). 如果假设被拒绝, 继续进行检验数目是 m_0+1. 继续直到假设接受或者直到 $\frac{1}{2}[(p-m)^2-p-m]\leqslant 0$. 最后甚至断定没有非平凡的因子模型适合. 可惜的是, 这个方法的误差的概率是未知的, 甚至渐近的误差的概率也是未知的.

14.3.3 估计量的渐近分布

极大似然估计量 $\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}$ 极大化平均集中对数似然函数 $L^*(\boldsymbol{C},\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*)$, 其由 (3) 式给出, 由 N 分成为 $\boldsymbol{\Sigma}^*=\boldsymbol{\Psi}^*+\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*'}$, 受 $\boldsymbol{\Lambda}^{*'}\boldsymbol{\Psi}^{*-1}\boldsymbol{\Lambda}^*$ 是对角的限制. 如果 $\boldsymbol{C}$ 是 $\boldsymbol{\Sigma}$ 的一致估计量 ("真实的" 协方差阵), 则在 $\boldsymbol{\Lambda},\boldsymbol{\Psi}$, 邻域概率一致地 $L^*(\boldsymbol{C},\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*)\to L^*(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}',\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*)$, $L^*(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}',\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*)$ 在 $\boldsymbol{\Psi}^*=\boldsymbol{\Psi}$ 和 $\boldsymbol{\Lambda}^*=\boldsymbol{\Lambda}$ 有唯一的极大值. 因为函数是连续的, 极大化 $L^*(\boldsymbol{C},\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*)$ 的 $\boldsymbol{\Lambda}^*,\boldsymbol{\Psi}^*$ 必须随机收敛到 $\boldsymbol{\Lambda},\boldsymbol{\Psi}$.

定理 14.3.1 如果 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Psi}$ 由 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的识别, 如果对角元是不同的和有次序的, 如果 $\boldsymbol{C}\xrightarrow{p}\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, 则 $\hat{\boldsymbol{\Psi}}\xrightarrow{p}\boldsymbol{\Psi}$ 和 $\hat{\boldsymbol{\Lambda}}\xrightarrow{p}\boldsymbol{\Lambda}$.

$\boldsymbol{C}\xrightarrow{p}\boldsymbol{\Sigma}$ 的一个充分条件是 $(f'\ U')'$ 的分布有有限的二阶矩.

估计量 $\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Psi}}$ 是方程 (10), (15) 的解, 要求 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的. 这些方程是多项式方程. $\hat{\boldsymbol{\Lambda}}$ 和 $\boldsymbol{\Psi}$ 的导数, 作为 $\boldsymbol{C}$ 的函数, 是连续的, 除非它们是无穷的. Anderson and Rubin(1956) 研究了导数是有穷时的条件并证明了下面的定理:

定理 14.3.2 令

$$(\theta_{ij})=\boldsymbol{\Theta}=\boldsymbol{\Psi}-\boldsymbol{\Lambda}(\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'. \tag{28}$$

如果 (θ_{ij}^2) 是非奇异的, 如果 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Psi}$ 被条件 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角的且对角元是不同的和有次序的所识别, 如果 $\boldsymbol{C}\xrightarrow{p}\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, 如果 $\sqrt{N}(\boldsymbol{C}-\boldsymbol{\Sigma})$ 有极限正态分布, 则 $\sqrt{N}(\hat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})$ 和 $\sqrt{N}(\hat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})$ 有极限正态分布.

举例来说, 如果 $(f'\ U')'$ 的分布有有穷四阶矩, 则 $\sqrt{N}(\boldsymbol{C}-\boldsymbol{\Sigma})$ 有极限分布.

这里很难得到 $\sqrt{N}(\hat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})$ 和 $\sqrt{N}(\hat{\boldsymbol{\Psi}}-\boldsymbol{\Psi})$ 的极限分布的协方差阵. Lawley(1953) 对未知的 $\boldsymbol{\Psi}$ 找到 $\sqrt{N}(\hat{\boldsymbol{\Lambda}}-\boldsymbol{\Lambda})$ 的合适的协方差, Lawley(1967) 在 $\boldsymbol{\Psi}$

是估计的情况下扩展了他的工作. [见 Lawley and Maxwell(1971).]Jennrich and Thayer(1973) 纠正了 Lawley 的工作中的一个错误.

在极限分布中, $\sqrt{N}(\hat{\psi}_{ii}-\psi_{ii})$ 和 $\sqrt{N}(\hat{\psi}_{jj}-\psi_{jj})$ 的协方差是

$$2\psi_{ii}^2\psi_{jj}^2\xi^{ij},\quad i,j=1,\cdots,p, \tag{29}$$

其中 $(\xi^{ij})=(\theta_{ij}^2)^{-1}$. 其他的协方差太复杂了, 不便在这里给出.

当渐近协方差很难给出样本的变异性时, 它们可以通过编程计算. 在那种情况下, 参数被它们的相容估计量取代.

14.3.4 极小距离法

极大似然的另一个形式是广义最小二乘. 估计量是极小化下式的 $\boldsymbol{\Psi}$ 和 $\boldsymbol{\Lambda}$ 的值,

$$\operatorname{tr}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{H}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{H}, \tag{30}$$

其中 $\boldsymbol{\Sigma}=\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$, $\boldsymbol{H}=\boldsymbol{\Sigma}^{-1}$ 或者某些 $\boldsymbol{\Sigma}^{-1}$ 的相容估计量. 当 $\boldsymbol{H}=\boldsymbol{\Sigma}^{-1}$ 时, 目标函数有形式

$$[\boldsymbol{c}-\boldsymbol{\sigma}(\boldsymbol{\Psi},\boldsymbol{\Lambda})]'[\operatorname{cov}\boldsymbol{c}]^{-1}[\boldsymbol{c}-\boldsymbol{\sigma}(\boldsymbol{\Psi},\boldsymbol{\Lambda})], \tag{31}$$

其中 $\boldsymbol{c}$ 表示 $\boldsymbol{C}$ 的排列在向量中的元, $\boldsymbol{\sigma}(\boldsymbol{\Psi},\boldsymbol{\Lambda})$ 是 $\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ 排列在相应的向量, $\operatorname{cov}\boldsymbol{c}$ 是在正态下的 $\boldsymbol{c}$ 的协方差阵 [Anderson(1973a)]. Jöreskog and Goldberger(1972) 用 $\boldsymbol{C}^{-1}$ 代替 $\boldsymbol{H}$, 极小化

$$\operatorname{tr}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{C}^{-1}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{C}^{-1}=\operatorname{tr}(\boldsymbol{I}-\boldsymbol{\Sigma}\boldsymbol{C}^{-1})^2. \tag{32}$$

令矩阵的元关于 $\boldsymbol{\Lambda}$ 的元的微商等于 0, 形成矩阵方程

$$\boldsymbol{C}^{-1}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{C}^{-1}\boldsymbol{\Lambda}=\boldsymbol{0}. \tag{33}$$

这个可以写成

$$\boldsymbol{\Lambda}=\boldsymbol{\Sigma}\boldsymbol{C}^{-1}\boldsymbol{\Lambda}. \tag{34}$$

左端乘 $\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}^{-1}$ 得到 (8) 式, 继而得到 (10) 式. 给定 $\boldsymbol{\Psi}$ 时 $\boldsymbol{\Lambda}$ 的估计量与极大似然估计量相同, 除了列的正规化. 这个方程由令 (32) 式关于 $\boldsymbol{\Psi}$ 的导数等于 0 得到

$$\operatorname{diag}\boldsymbol{C}^{-1}[(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}')-\boldsymbol{C}]\boldsymbol{C}^{-1}=\operatorname{diag}\boldsymbol{0}. \tag{35}$$

一个可选择的方案是极小化

$$\frac{1}{2}\operatorname{tr}\{(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}')^{-1}[\boldsymbol{C}-(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}')]\}^2. \tag{36}$$

这个导出了 (8) 或者 (10), 以及

$$\operatorname{diag}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}=\operatorname{diag}\boldsymbol{0}. \tag{37}$$

Browne(1974) 证明了 $\boldsymbol{\Psi}$ 的广义最小二乘估计量和极大似然估计量有相同的渐近分布. Dahm and Fuller(1981) 证明了如果 (31) 式中的 $\operatorname{cov}\boldsymbol{c}$ 被收敛到 $\operatorname{cov}\boldsymbol{c}$ 的矩阵替代, $\boldsymbol{\Psi},\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Phi}$ 取决于一些参数, 则它和极大似然的渐近分布是相同的.

14.3.5 同主成分分析的关系

Hotelling(1933) 提出的极大似然和主成分分析的关系是什么? 如同第 11 章解释的, 样本主成分的向量是 $\boldsymbol{B}'\boldsymbol{X}$ 的正交变换, 其中 $\boldsymbol{B}$ 的列是 $\boldsymbol{C}$ 的特征向量, 由 $\boldsymbol{B}'\boldsymbol{B}=\boldsymbol{I}$ 正规化. 则

$$\boldsymbol{C}=\boldsymbol{B}\boldsymbol{T}\boldsymbol{B}'=\sum_{i=1}^{p}\boldsymbol{b}_i t_i \boldsymbol{b}_i', \tag{38}$$

其中 $\boldsymbol{T}$ 是正交矩阵, 对角元 $t_1,\cdots,t_p$ 是 $\boldsymbol{C}$ 的特征根. 如果 $t_{m+1},\cdots,t_p$ 是较小的, $\boldsymbol{C}$ 可以由下式近似,

$$\boldsymbol{B}_1\boldsymbol{T}_1\boldsymbol{B}_i'=\sum_{i=1}^{m}\boldsymbol{b}_i t_i \boldsymbol{b}_i', \tag{39}$$

其中 $\boldsymbol{T}_1$ 是对角矩阵, 有对角元 $t_1,\cdots,t_m$, $\boldsymbol{X}$ 由下式近似,

$$\boldsymbol{B}_1\boldsymbol{B}_1'\boldsymbol{X}=\sum_{i=1}^{m}\boldsymbol{b}_i(\boldsymbol{b}_i'\boldsymbol{X}). \tag{40}$$

则 $\boldsymbol{X}$ 与近似 (40) 之差的样本协方差是

$$\boldsymbol{X}-\boldsymbol{B}_1\boldsymbol{B}_1'\boldsymbol{X}=\boldsymbol{B}_2\boldsymbol{B}_2'\boldsymbol{X} \tag{41}$$

的样本协方差, 即 $\boldsymbol{B}_2\boldsymbol{T}_2\boldsymbol{B}_2'=\sum_{i=m+1}^{p}\boldsymbol{b}_i t_i \boldsymbol{b}_i'$, 分量的方差的和是 $\sum_{i=m+1}^{p}t_i$. 这里 $\boldsymbol{T}_2$ 是以 $t_{m+1},\cdots,t_p$ 为对角元的对角矩阵.

这个分析是就一些度量的公共单位而言的. 前 m 个元 "解释" "方差" 的一个较大的比例, $\operatorname{tr}\boldsymbol{C}$. 当度量单位不相同时 (比如, 当单位是任意的), 通常将度量 (样本) 标准化为方差是 1. 尽管如此, 主成分没有用方差来解释.

主成主分析和因子分析的另一个区别是, 前者没有从系统部分分离出误差. 但是这个缺点很容易补救. Thomson (1934) 对因子分析模型提出了下面的估计方法. 从 $\boldsymbol{C}$ 中减去一个对角矩阵 $\boldsymbol{\Psi}$, 主成主分析在 $\boldsymbol{C}-\boldsymbol{\Psi}$ 上实施. 尽管如此, $\boldsymbol{\Psi}$ 是已定的, 所以 $\boldsymbol{C}-\boldsymbol{\Psi}$ 接近于秩 m . 方程是

$$(\boldsymbol{C}-\boldsymbol{\Psi})\boldsymbol{\Lambda}=\boldsymbol{\Lambda}\boldsymbol{L}, \tag{42}$$

$$\operatorname{diag}(\boldsymbol{\Psi}+\boldsymbol{\Lambda}\boldsymbol{\Lambda}')=\operatorname{diag}C, \tag{43}$$

$$\boldsymbol{\Lambda}'\boldsymbol{\Lambda}=\boldsymbol{L}\,(\text{对角的}). \tag{44}$$

最后的等式是正规化和除去 $\boldsymbol{\Lambda}$ 的不确定性. 这个方法允许误差项, 但是取决于度量的单位. 估计量是相容的但是在通常的因子分析模型中不是 (渐近地) 有效的.

14.3.6 形心方法

在高速计算机出现之前, 由于计算的容易. 形心方法几乎独占, 由于历史的缘故我们给出这种方法的梗概. 令 $\boldsymbol{R}^*$ 是相关约化矩阵, 即矩阵的元素是 $r_{ij}, i\neq j$, 还有 $1-\hat{\psi}_{ii}^*$, 其中 $\hat{\psi}_{ii}^*$ 是一个误差方差在标准差单位下的初始估计量. Thomson 的主成主近似首先找到 $\boldsymbol{R}_0=\boldsymbol{R}^*$ 的对应于第 m 个较大的特征根的第 m 个特征向量.

像第 11 章中简述的, 一个计算方法包括从第一个向量的初始估计开始, 设为 $\boldsymbol{x}^{(0)}$, 计算 $\boldsymbol{x}^{(1)} = \boldsymbol{R}_0\boldsymbol{x}^{(0)}$, 逐步迭代. 第 r 步 $\boldsymbol{x}^{(r)}$ 近似地是 $\gamma_1\boldsymbol{x}^{(r-1)}$, 其中 γ_1 是最大的根, 并且有 $\boldsymbol{x}^{(r)\prime}\boldsymbol{x}^{(r)} \sim \gamma_1^2\boldsymbol{x}^{(r-1)\prime}\boldsymbol{x}^{(r-1)}$. 则 $y_1 = \boldsymbol{x}^{(r)}/\sqrt{\gamma_1\boldsymbol{x}^{(r-1)\prime}\boldsymbol{x}^{(r-1)}}$ 近似地是第一个正规化特征向量, 所以 $\boldsymbol{y}_1'\boldsymbol{y}_1 = \gamma_1$. 为了得到第二个向量, 对 $\boldsymbol{R}_1 = \boldsymbol{R}^* - \boldsymbol{y}_1\boldsymbol{y}_1'$ 应用相同的方法.

形心方法可以看作主成分方法的一个非常粗糙的近似. 心理学测试的相关阵通常由正的元素组成, 第一个特征向量有所有正的分量, 常常有大约相同的值. 形心方法用 $\varepsilon = (1, \cdots, 1)'$ 作为第一个向量的初始估计. 则 $\boldsymbol{R}^*\varepsilon = \boldsymbol{x}^{(1)}$ 是第一次迭代, 应该是对第一个特征向量的近似. 第一个特征根的一个近似是 $\varepsilon'\boldsymbol{R}^*\varepsilon/\varepsilon'\varepsilon$. 则 $\boldsymbol{y}_1 = \boldsymbol{x}^{(1)}/\sqrt{\varepsilon'\boldsymbol{R}^*\varepsilon}$ 是一个当 $\boldsymbol{R}^*$ 正规化后有平方长度 γ_1 时的第一个特征向量的近似. 可以在加法机或者台式计算器上实施计算, 因为 $\boldsymbol{R}^*\varepsilon$ 相当于增加行, 并且 $\varepsilon'\boldsymbol{R}^*\varepsilon$ 是那些行的总和.

第二个特征向量同第一个是正交的. 同 ε 正交的向量 ε^*, 它由 $p/2$ 个 1 和 $p/2$ 个 -1 组成. 则 $\boldsymbol{R}_1\varepsilon^* = \boldsymbol{x}_2$ 是对第二个特征向量的近似, $\varepsilon^{*\prime}\boldsymbol{R}_1\varepsilon^*/\varepsilon^{*\prime}\varepsilon^*$ 近似了第二个特征根. 这些操作包括变换 $\boldsymbol{R}_1$ 的元素的正负和增加元素. 选择 ε^* 中的 -1 的位置使得极大化 $\varepsilon^{*\prime}\boldsymbol{R}_1\varepsilon^*$. 方法可以继续.

14.4 不变因子的估计

令 $\boldsymbol{x}_\alpha = (x_{1\alpha}, \cdots, x_{p\alpha})'$ 是由下式给出的 $\boldsymbol{X}_\alpha$ 的一组观测,

$$\boldsymbol{X}_\alpha = \boldsymbol{\Lambda}\boldsymbol{f}_\alpha + \boldsymbol{\mu} + \boldsymbol{U}_\alpha, \tag{1}$$

$\boldsymbol{f}_\alpha(\alpha = 1, \cdots, N)$ 是一个非随机向量 (一个附加参数), 满足 $\sum_{\alpha=1}^N \boldsymbol{f}_\alpha = \boldsymbol{0}$. 似然函数是

$$L = \frac{1}{\left[(2\pi)^p \prod_{i=1}^p \psi_{ii}\right]^{N/2}} \prod_{i=1}^p \exp\left\{-\frac{1}{2}\sum_{\alpha=1}^N \frac{(x_{i\alpha} - \mu_i - \sum_{j=1}^m \lambda_{ij}f_{j\alpha})^2}{\psi_{ii}}\right\}. \tag{2}$$

这个似然函数没有极大值. 为了证明这个事实, 令 $\mu_1 = 0, \lambda_{11} = 1, \lambda_{1j} = 0 (j \neq 1), f_{1\alpha} = x_{1\alpha}$. 则 $x_{1\alpha} - \mu_1 - \sum_{j=1}^m \lambda_{ij}f_{j\alpha} = 0$, ψ_{11} 并没有在指数中出现而是只在常数出现. 当 $\psi_{11} \to 0, L \to \infty$. 因此似然没有极大值, 极大似然估计量不存在 [Anderson and Rubin (1956)]. Lawley(1941) 令似然的偏导数等于 0, 但是 Solari(1969) 证明了解只是一个平稳值, 不是极大值.

由于极大似然估计量在不变因子的情况下不存在, 那么可以用什么估计方法呢? 一个可能是用适合随机因子的极大似然方法. Fuller, Pantula, and Amemiya (1982) 对 0 识别的情形证明了极大似然估计量的渐近正态分布在随机情况时和不变因子时相同但这是由 Anderson and Rubin(1956) 提出的.

在正态性下的样本协方差阵有非中心 Wishart 分布 [Anderson(1946a)], 它取决于 $\boldsymbol{\Psi}, \boldsymbol{\Lambda\Phi\Lambda}', N-1$. Anderson and Rubin(1956) 论证了极大化这个似然函数. 但是, 其中一个方程很难求解. 估计量渐近等价于随机-因子情况下的极大似然估计量.

14.5 因子的解释和变换

14.5.1 解释

识别条件限制 $\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ 是对角或者 $\boldsymbol{\Lambda}$ 的前 m 行是 $\boldsymbol{I}_m$, 可能使极大似然估计量的计算更简单, 但是因子得分向量的分量可能没有任何本质的意义. 我们在 14.2 节看到, 0 系数对因子的可能的意义是这个因子不影响某种测试. 类似地, 较大的因子载荷可能帮助解释一个因子. 比如, 语言能力的系数, 在语言能力强时应该在测试中比较大.

在心理学中, 每个变量或者因子通常有其自然的正向: 测试中的正确答案越多, 由因子表示出的能力越强. 通常期望能力越强表现越好即因子载荷如果非 0 则应该是正的. 因此, 严格地说, 由于解释的缘故, 应该寻找的因子载荷或者为 0 或者是较大的正数.

14.5.2 变换

基于某些任意的识别条件 (包括 $\boldsymbol{\Phi}=\boldsymbol{I}$) 的极大似然估计量是 $\hat{\boldsymbol{\Lambda}}$ 和 $\hat{\boldsymbol{\Phi}}$. 我们考虑变换

$$\hat{\boldsymbol{\Lambda}}^* = \hat{\boldsymbol{\Lambda}}\boldsymbol{P}, \quad \hat{\boldsymbol{\Phi}}^* = \boldsymbol{P}^{-1}(\boldsymbol{P}^{-1})' = (\boldsymbol{P}'\boldsymbol{P})^{-1}. \tag{1}$$

如果因子是正交的, 则 $\boldsymbol{\Phi}^*=\boldsymbol{I}$ 和 $\boldsymbol{P}$ 是正交的. 如果允许因子是偏斜的, 那么 $\boldsymbol{P}$ 可以是任意非奇异矩阵, $\hat{\boldsymbol{\Phi}}^*$ 是任意正定矩阵.

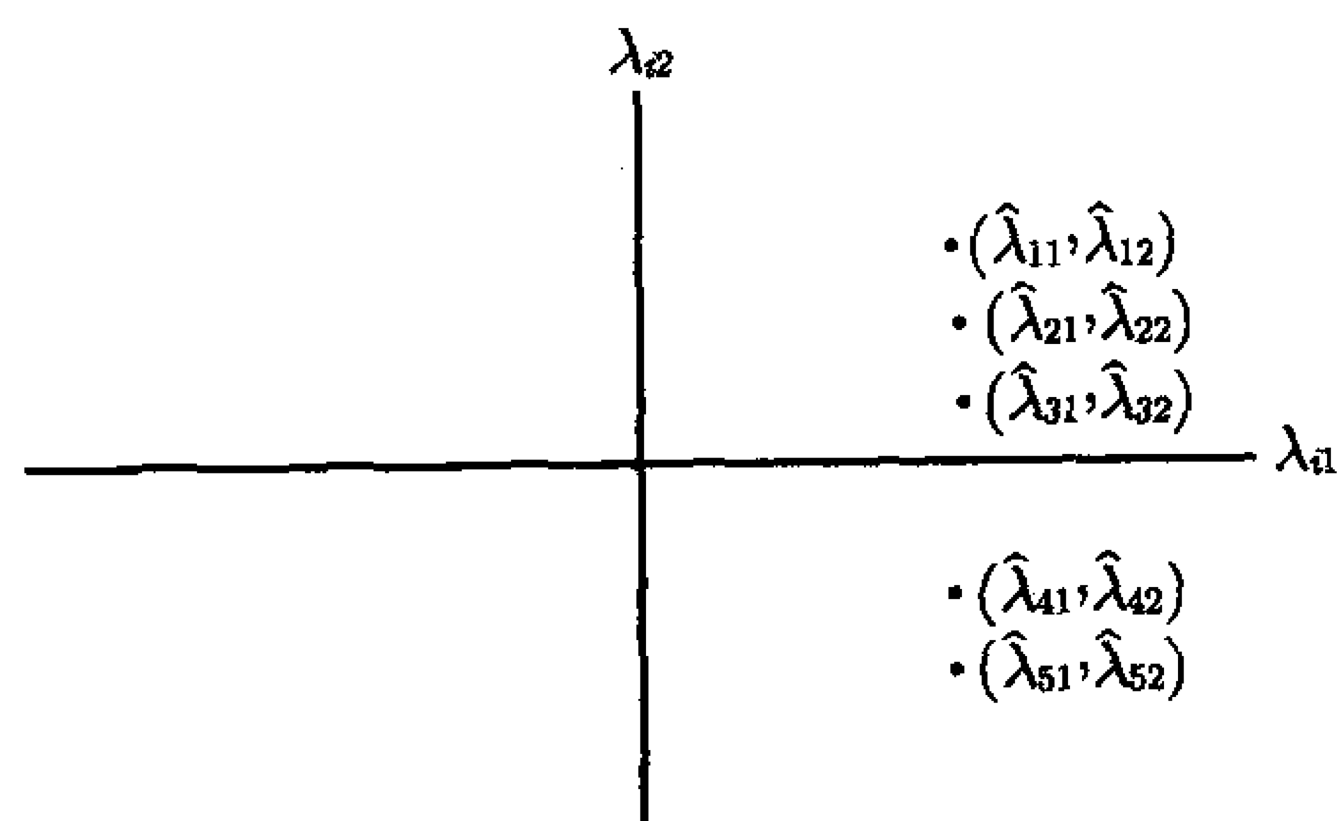


图 14.1 $\hat{\boldsymbol{\Lambda}}$ 的行

$\hat{\boldsymbol{\Lambda}}$ 的行可以在 m 维的空间中画出. 图 14.1 是 5×2 矩阵 $\hat{\boldsymbol{\Lambda}}$ 的行的图. 坐标涉及因子, 点涉及测试. 如果要求 $\boldsymbol{\Phi}^*$ 为 $\boldsymbol{I}_m$, 我们寻求这个空间中坐标轴的旋转. 图示的这个例子, 45° 旋转将会把所有的点放入正的象限, 即 $\lambda_{ij}^*\geqslant 0$. 新坐标系之一使得前三个点变大, 其他两个点变小, 其他的坐标将会使得前三个点小后两个点大. 第一个因子代表前三个测试的公共因素, 第二个因子代表最后两个测试的公共因素.

如果 $m>2$, 一个一般的旋转可以由一系列手工的两维旋转近似.

如果不要求 $\boldsymbol{\Phi}^*$ 是 $\boldsymbol{I}_m$，变换 $\boldsymbol{P}$ 是非奇异的. 如果 $\boldsymbol{\Lambda}$ 的第 j 列的正规化是 $\lambda_{i(j),j}=1$, 则

$$1=\hat{\lambda}^*_{i(j),j}=\sum_{k=1}^{m}\hat{\lambda}_{i(j),k}p_{kj}, \tag{2}$$

$\boldsymbol{P}$ 的每列都满足这样的限制. 如果正规化是 $\phi_{jj}=1$, 则

$$1=\phi^*_{jj}=\sum_{k=1}^{m}(p^{jk})^2, \tag{3}$$

其中 $(p^{jk})=\boldsymbol{P}^{-1}$.

基于极大化一个目标函数的计算方法有很多, 我们仅讨论由 Kaiser(1958) 提出的方差极大法, 这是对成对的因子实施的. Horst(1965) 的第 18 章将方法扩展到对所有的因子同时适用. 一个修正的准则是

$$\sum_{j=1}^{m}\sum_{i=1}^{p}\left(\lambda_{ij}^{*2}-\frac{\sum_{h=1}^{p}\lambda_{hj}^{*2}}{p}\right)^2=\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\lambda_{ij}^{*4}-\frac{(\sum_{h=1}^{p}\lambda_{hj}^{*2})^2}{p}\right], \tag{4}$$

它同变换因子载荷的平方的列方差的和成比例. 选择正交矩阵 $\boldsymbol{P}$ 使得极大化 (4) 式. 方向趋向极大化 λ_{ij}^{*2} 在列内的散布. 由于 $\lambda_{ij}^{*2}\geqslant 0$, 将获得一些较大的载荷和一些在 0 附近的载荷. Kaiser 的初始准则是 (4) 式, 将 λ_{ij}^{*2} 用 $\lambda_{ij}^{*2}/\sum_{h=1}^{m}\lambda_{ih}^{*2}$ 取代.

Lawley and Maxwell(1971) 描述了其他准则. 其中一个是度量类似于预先确定由 1 和 0 组成的矩阵.

14.5.3 正交因子与偏斜因子

对于正交因子, 总体中或者样本中分量不相关取决于因子是随机的还是不变的. 不相关因子得分的思想已经讨论过. 一些心理学家宣称, 如果更多地考虑因子得分而不是测试得分, 因子得分的正交性是至关重要的. 关于这点心理学家有相当多的辩论. 另一方面, Thurstone(1947), 第 vii 页, 讲到 “看起来不必要求在一般总体中智力特征不相关, 就如同要求在一般总体中身高和体重不相关一样”.

如我们看到的, 给定的两个矩阵 $\boldsymbol{\Lambda},\boldsymbol{\Phi}$, 等价地由 $\boldsymbol{\Lambda P},\boldsymbol{P}^{-1}\boldsymbol{\Phi P}'^{-1}$ 对非奇异的 $\boldsymbol{P}$ 给出. 这两个矩阵可能 (比如给定 $\boldsymbol{\Lambda},\boldsymbol{\Phi}$ 后的 $\boldsymbol{P}$) 具有最有意义的解释 (就测试的目标事件而言). 简单结构的思想是有 0 因子载荷, 某种形式可以给出分量因子得分的意义而不考虑矩量矩阵. 允许 $\boldsymbol{\Phi}$ 是一个任意的正定矩阵, 相当于允许在 $\boldsymbol{\Lambda}$ 中有更多的 0.

选择变换或者识别条件的另一个考虑是自主的, 或者永久性的, 或者关于某种变化有不变性. 比如, 如果选择一个总体的成分将会发生什么? 在智力测试的情况下, 假设作了一个选择, 比如大学招收高中高年级学生, 假设有某种主要的技能. 可以设想非观测因子得分 $\boldsymbol{f}$ 和没被选择影响的观测测试得分 $\boldsymbol{x}$ 的关系, 即因子载荷矩阵 $\boldsymbol{\Lambda}$ 没变. 可以认为误差的方差 (特殊因子), $\boldsymbol{\Psi}$ 的对角元, 关于由选择是不变的, 因为误差和因子是不相关的 (主要的技能).

假设有一个真实模型, $\mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}$, 调查者应用准许他发现它的识别条件. 下一步, 假设有一个选择, 它产生一个新的因子得分总体使得他们的协方差阵是 $\mathbf{\Phi}^*$. 当调查者分析新的观测协方差阵 $\mathbf{\Psi}+\mathbf{\Lambda}\mathbf{\Phi}^*\mathbf{\Lambda}'$ 时, 将会再得到 $\mathbf{\Lambda}$ 吗? 如果识别条件的一部分是因子矩量矩阵是 $\boldsymbol{I}$, 然后他将得到一个不同的因子载荷矩阵. 另一方面, 如果识别条件完全在因子载荷 (给定的 0 和 1), 从分析可知因子载荷矩阵和前面是相同.

一个相同的考虑与比较两个总体有关. 很合理地考虑 $\mathbf{\Psi}_1=\mathbf{\Psi}_2, \mathbf{\Lambda}_1=\mathbf{\Lambda}_2$, 但是 $\mathbf{\Phi}_1 \neq \mathbf{\Phi}_2$. 为了检验假设 $\mathbf{\Phi}_1=\mathbf{\Phi}_2$, 想要利用同 $\mathbf{\Lambda}_1=\mathbf{\Lambda}_2$ 一致的识别条件 (而不是 $\mathbf{\Lambda}_1=\mathbf{\Lambda}_2\boldsymbol{C}$). 条件加在因子载荷上.

如果加上更多的测试 (或者删除) 将会怎样? 除了观测 $\boldsymbol{X}=\mathbf{\Lambda}\boldsymbol{f}+\boldsymbol{\mu}+\boldsymbol{U}$ 之外, 假设还有观测 $\boldsymbol{X}^*=\mathbf{\Lambda}^*\boldsymbol{f}+\boldsymbol{\mu}^*+\boldsymbol{U}^*$, 其中 $\boldsymbol{U}^*$ 是同 $\boldsymbol{U}$ 不相关的. 由于公共因子 $\boldsymbol{f}$ 是未改变的, $\mathbf{\Phi}$ 是不变的. 但是, (任意的) 条件 $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ 是对角的却变化了, 用这种条件将会产生 $(\mathbf{\Lambda}'\,\mathbf{\Lambda}^{*'})$ 的旋转.

14.6　指定零识别的估计

我们现在考虑当 $\mathbf{\Phi}$ 是无约束的和 $\mathbf{\Lambda}$ 是由指定的 0 和 1 识别时, $\mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Phi}$ 的估计. 我们假设 $\mathbf{\Lambda}$ 的每列有至少 $m+1$ 个 0 在指定位置, 由 $\mathbf{\Lambda}$ 的行组成的子矩阵秩为 $m-1$, $\mathbf{\Lambda}$ 在给定的列有指定的 0. (见 14.2.2 节.) 我们进一步假设 $\mathbf{\Lambda}$ 的每列有 1 在指定的位置, 或者, $\mathbf{\Phi}$ 相应于那个列的对角元是 1. 则模型是给定的.

似然函数由 14.3 节的 (1) 式给出. 令似然函数的导数等于零, 就是

$$\operatorname{diag}\mathbf{\Sigma}^{-1}[\boldsymbol{C}-(\mathbf{\Psi}+\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}')]\mathbf{\Sigma}^{-1}=\operatorname{diag}\mathbf{0}, \tag{1}$$

$$\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}[\boldsymbol{C}-(\mathbf{\Psi}+\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}')]\mathbf{\Sigma}^{-1}\mathbf{\Lambda}=\mathbf{0}, \tag{2}$$

$\mathbf{\Phi}$ 中的位置并没有指定, 还有

$$\mathbf{\Sigma}^{-1}[\boldsymbol{C}-(\mathbf{\Psi}+\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}')]\mathbf{\Sigma}^{-1}\mathbf{\Lambda}=\mathbf{0}, \tag{3}$$

$\mathbf{\Lambda}$ 中的位置并未指定, 其中

$$\mathbf{\Sigma}=\mathbf{\Psi}+\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}. \tag{4}$$

这些方程不能像 14.3.1 节中那样简化, 因为 (3) 式仅对 $\mathbf{\Lambda}$ 中未指定的位置成立. 因此它不能在左端乘 $\mathbf{\Sigma}$. [见 Howe(1955),Anderson and Rubin(1956), Lawley(1958)].

这些方程对计算并无用处. 但是似然函数, 可以在数值上极大化.

前面提到, 度量单位的变化, $\boldsymbol{X}^*=\boldsymbol{D}\boldsymbol{X}$, 导致了参数 $\mathbf{\Lambda}$ 和 $\mathbf{\Psi}$ 相应的变化, 如果在 $\mathbf{\Lambda}$ 指定的位置由零 0 识别, 并且由 $\phi_{jj}=1(j=1,\cdots,m)$ 正规化. 很容易证明导数方程 (1),(2),(3), (4) 以相应的方式改变.

Anderson and Amemiya(1988a) 在一般条件下得到了估计量的渐近分布. 并没有要求观测的正态性. 见 Anderson and Amemiya(1988b).

14.7 因子得分的估计

我们常常对研究组的个体的因子得分感兴趣. 在非随机因子的模型中因子得分是特征化个体的非主要参数. 正如我们看到的 (14.4 节), 参数 $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\mu}, \boldsymbol{f}_1, \cdots, \boldsymbol{f}_N)$ 的极大似然估计量不存在. 我们因此基于结构参数 $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\mu})$ 是未知时研究因子得分的估计.

当把 $\boldsymbol{f}_\alpha$ 是视为一个附加参数时, $\boldsymbol{x}_\alpha - \boldsymbol{\mu}$ 是一个来自有均值 $\boldsymbol{\Lambda f}_\alpha$ 和协方差阵 $\boldsymbol{\Psi}$ 的分布的观测. $\boldsymbol{f}_\alpha$ 的加权最小二乘估计量是

$$\begin{aligned}\hat{\boldsymbol{f}}_\alpha &= (\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}) \\ &= \boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}),\end{aligned} \tag{1}$$

其中 $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$(不必是对角的). 这个估计量是无偏的, 由通常的最小二乘理论 [Bartlett(1937b),(1938)] 可知, 它的协方差阵是

$$\mathrm{E}(\hat{\boldsymbol{f}}_\alpha - \boldsymbol{f}_\alpha)(\hat{\boldsymbol{f}}_\alpha - \boldsymbol{f}_\alpha)' = (\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1} = \boldsymbol{\Gamma}^{-1}. \tag{2}$$

它是 $\boldsymbol{f}_\alpha$ 的极小方差线性无偏估计量. 如果 $\boldsymbol{x}_\alpha$ 是正态的, 则估计量也是极大似然的.

当 $\boldsymbol{f}_\alpha$ 是随机的 [Thomson(1951)], 我们假设 $\boldsymbol{X}_\alpha$ 和 $\boldsymbol{f}_\alpha$ 有联合正态分布, 有均值向量 $(\boldsymbol{\mu}', \boldsymbol{0}')$ 和协方差阵

$$\mathrm{Cov}\begin{pmatrix} X \\ \boldsymbol{f} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Psi} + \boldsymbol{\Lambda\Phi\Lambda}' & \boldsymbol{\Lambda\Phi} \\ \boldsymbol{\Phi\Lambda}' & \boldsymbol{\Phi} \end{pmatrix}. \tag{3}$$

则 f 对 $\boldsymbol{X}$ 的回归是 (2.5 节)

$$\begin{aligned}\mathrm{E}(f|\boldsymbol{X}) &= \boldsymbol{\Phi\Lambda}'(\boldsymbol{\Phi} + \boldsymbol{\Lambda\Phi\Lambda}')^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \\ &= \boldsymbol{\Phi}(\boldsymbol{\Phi} + \boldsymbol{\Phi\Gamma\Phi})^{-1}\boldsymbol{\Phi\Lambda}'\boldsymbol{\Phi}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}).\end{aligned} \tag{4}$$

$\boldsymbol{f}_\alpha$ 的估计量或预估是

$$\hat{\boldsymbol{f}}_\alpha^* = \boldsymbol{\Phi}(\boldsymbol{\Phi} + \boldsymbol{\Phi\Gamma\Phi})^{-1}\boldsymbol{\Phi\Lambda}'\boldsymbol{\Phi}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}). \tag{5}$$

如果 $\boldsymbol{\Phi} = \boldsymbol{I}$, 预估是

$$\hat{\boldsymbol{f}}_\alpha^* = (\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu}). \tag{6}$$

当 $\boldsymbol{\Gamma}$ 也是对角时, (6) 式的第 j 个元是 $\gamma_j/(1+\gamma_j)$ 倍的 (1) 式的第 j 个元. 给定 $\boldsymbol{f}_\alpha$ 时 $(\boldsymbol{\Phi} = \boldsymbol{I})$, $\boldsymbol{x}_\alpha$ 的条件分布是

$$\mathrm{E}(\hat{\boldsymbol{f}}_\alpha^*|\boldsymbol{f}_\alpha) = (\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}\boldsymbol{f}_\alpha, \tag{7}$$

$$\mathrm{Cov}(\hat{\boldsymbol{f}}_\alpha^*|\boldsymbol{f}_\alpha) = (\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}(\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}, \tag{8}$$

$$\mathrm{E}[(\hat{\boldsymbol{f}}_\alpha^* - \boldsymbol{f}_\alpha)(\hat{\boldsymbol{f}}_\alpha^* - \boldsymbol{f}_\alpha)'|\boldsymbol{f}_\alpha] = (\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}(\boldsymbol{\Gamma} + \boldsymbol{f}_\alpha\boldsymbol{f}_\alpha')(\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}, \tag{9}$$

$$\mathrm{E}(\hat{\boldsymbol{f}}_\alpha^* - \boldsymbol{f}_\alpha)(\hat{\boldsymbol{f}}_\alpha^* - \boldsymbol{f}_\alpha)' = (\boldsymbol{I} + \boldsymbol{\Gamma})^{-1}. \tag{10}$$

最后这个矩阵描述了均方误差, 它比 (2) 式描述的无偏估计量小. 估计量 (5) 或者 (6) 是一个贝叶斯估计量, 比把 $\boldsymbol{f}_\alpha$ 当作随机时更恰当.

习 题

14.1 (14.2 节) 由 0 的识别. 令

$$\boldsymbol{\Lambda}=\begin{pmatrix} \mathbf{0} & \boldsymbol{\Lambda}^{(1)} \\ \boldsymbol{\lambda}_{(1)} & \boldsymbol{\Lambda}_{(1)} \end{pmatrix}, \qquad \boldsymbol{C}=\begin{pmatrix} c_{11} & \boldsymbol{c}_{12} \\ \boldsymbol{c}_{21} & \boldsymbol{C}_{22} \end{pmatrix},$$

其中 $\boldsymbol{C}$ 是非奇异的. 证明

$$\boldsymbol{\Lambda C}=\begin{pmatrix} \mathbf{0} & \boldsymbol{\Lambda}^{*(1)} \\ \boldsymbol{\lambda}^*_{(1)} & \boldsymbol{\Lambda}^*_{(1)} \end{pmatrix}$$

意味有

$$\boldsymbol{C}=\begin{pmatrix} c_{11} & \boldsymbol{c}_{12} \\ \mathbf{0} & \boldsymbol{C}_{22} \end{pmatrix}$$

当且仅当 $\boldsymbol{\Lambda}^{(1)}$ 的秩是 $m-1$.

14.2 (14.3 节) 对 $p=3, m=1, \boldsymbol{\Lambda}=\lambda$, 证明 $|\theta_{ij}^2|=\prod_{i=1}^3(\lambda_i^2/\psi_{ii})$.

14.3 (14.3 节) EM 算法

(a) 如果 $\boldsymbol{f}$ 和 $\boldsymbol{U}$ 是正态的, $\boldsymbol{f}$ 和 $\boldsymbol{X}$ 是观测, 证明基于 $(\boldsymbol{x}_1, \boldsymbol{f}_1), \cdots, (\boldsymbol{x}_N, \boldsymbol{f}_N)$ 的似然函数是

$$\prod_{\alpha=1}^{N}\left\{\frac{1}{(2\pi)^{\frac{1}{2}p}\prod_{i=1}^{p}\psi_{ii}}\exp\left[-\frac{1}{2}\sum_{i=1}^{p}\frac{(x_{i\alpha}-\mu_i-\sum_{j=1}^{m}\lambda_{ij}f_{j\alpha})^2}{\psi_{ii}}\right]\right.$$
$$\left.\cdot\frac{1}{(2\pi)^{\frac{1}{2}m}|\boldsymbol{\Phi}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\boldsymbol{f}'_\alpha\boldsymbol{\Phi}^{-1}\boldsymbol{f}_\alpha\right]\right\}.$$

(b) 证明: 当因子得分被列为数据时, 统计量的充分集是 $\bar{\boldsymbol{x}}, \bar{\boldsymbol{f}}, \boldsymbol{C}_{xx}=\boldsymbol{C}$,

$$\boldsymbol{C}_{xf}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha-\bar{\boldsymbol{x}})(\boldsymbol{f}_\alpha-\bar{\boldsymbol{f}})',$$

$$\boldsymbol{C}_{ff}=\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{f}_\alpha-\bar{\boldsymbol{f}})(\boldsymbol{f}_\alpha-\bar{\boldsymbol{f}})'.$$

(c) 证明: 给定 $\boldsymbol{X}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N), \boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$ 的条件下, (b) 中的协方差的条件期望是

$$\boldsymbol{C}^*_{xx}=\mathrm{E}(\boldsymbol{C}_{xx}|\boldsymbol{X},\boldsymbol{\Lambda},\boldsymbol{\Phi},\boldsymbol{\Psi})=\boldsymbol{C}_{xx},$$
$$\boldsymbol{C}^*_{xf}=\mathrm{E}(\boldsymbol{C}_{xf}|\boldsymbol{X},\boldsymbol{\Lambda},\boldsymbol{\Phi},\boldsymbol{\Psi})=\boldsymbol{C}_{xx}(\boldsymbol{\Psi}+\boldsymbol{\Lambda\Phi\Lambda}')^{-1}\boldsymbol{\Lambda\Phi},$$
$$\boldsymbol{C}^*_{ff}=\mathrm{E}(\boldsymbol{C}_{ff}|\boldsymbol{X},\boldsymbol{\Lambda},\boldsymbol{\Phi},\boldsymbol{\Psi})=\boldsymbol{\Phi\Lambda}'(\boldsymbol{\Psi}+\boldsymbol{\Lambda\Phi\Lambda}')^{-1}\boldsymbol{C}_{xx}(\boldsymbol{\Psi}+\boldsymbol{\Lambda\Phi\Lambda}')^{-1}\boldsymbol{\Lambda\Phi}$$
$$+\boldsymbol{\Phi}-\boldsymbol{\Phi\Lambda}'(\boldsymbol{\Psi}+\boldsymbol{\Lambda\Phi\Lambda}')^{-1}\boldsymbol{\Lambda\Phi}.$$

(d) 证明: 给定 $\boldsymbol{\Phi}=\boldsymbol{I}$ 时, $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Psi}$ 的极大似然估计量是

$$\hat{\boldsymbol{\Lambda}}=\boldsymbol{C}^*_{xf}\boldsymbol{C}^{*-1}_{ff},$$

$$\hat{\boldsymbol{\Psi}}=\boldsymbol{C}^*_{xx}-\boldsymbol{C}^*_{xf}\boldsymbol{C}^{*-1}_{ff}\boldsymbol{C}^{*'}_{xf}.$$

第 15 章　相依性模式, 图模型

15.1　引　　言

多元统计分析中的一个重点是, 对一个对象或个体的若干测量可能是相关的, 本书讨论的各种方法都考虑到这种相依性. 两个变量之间关联性的量可以用它们的 (Pearson) 相关来度量 (对称度量); 一个变量与一组变量之间的关系可以用一种多重相关来量化; 一组变量与另一组变量的联系可以通过第 9 章讲的独立性准则或者典型相关的方法来研究. 类似的度量可以用于条件分布的情形. 另一类相依性 (非对称的) 用回归系数和有关的度量来刻画. 在这章, 我们研究涉及几类相依性或者更复杂的相依性模式的模型.

统计中的*图模型*是指一种形象的图表, 在这种图表中, 那些可观测的变量用通过边连接的点 (*顶点*或*结点*) 和一个相联系的概率分布族来识别, 这种概率分布族满足由形象的模式所确定的某种独立性. 边可以是*无向的*(画为线段) 或者*有向的*(画为带箭头的线). 无向的边还分对称的相依性和独立性, 而有向的边可以反映一种可能的行动方向或者时间的序列. 这种独立性来自对客观事物认识的一种先验知识, 或者来自这些或另外的数据. 图展示的优点是容易理解, 特别是对复杂模式, 容易启示专家意见, 而且容易进行概率比较.

这种图表示法的使用至少要追朔到遗传学家 Sewall Wright (1921), (1934) 的工作, 他用的是"路径分析"术语. 关于图模型, 一种精细的代数理论已经发展起来. 独立性的描述化为要确定的参数个数. 这种独立性中的某些就是大家已知的Markov *性* (马氏性). 例如, 在 Markov 过程 (或一阶) 的时间序列分析中, 当现在为给定时, 认为过程的将来与过去是独立的; 在这样一个模型中, 过去的变量与将来变量之间的相关由当前的变量和最近的将来变量之间的相关所决定. 这种思想可用几种方法来解释.

与一个给定的图表相联系的概率分布族依赖于用图表示的分布性质. 由无向边组成的图表 (视作为*无向图*) 的性质将在 15.2 节里描述; 完全由有向边 (视作为*有向图*) 组成的图表的性质在 15.3 中描述; 而由这两种类型的边组成的图表的性质在 15.4 中讲述. 统计推断的方法将在 15.5 节里给出.

在本章, 我们假设变量具有联合非奇异的正态分布; 因此, 一个模型是借助协方差阵和它的逆以及它们的函数来刻画的. 这个假设意味着变量是定量的而且有

正的密度. 图模型的数学可以用于离散的变量 (列联表) 和非正态定量变量, 但是我们将不会讨论需要包含它们的这种理论.

继 Wright 的开创性工作后, 出现了相当多的社会科学文献. 有关这方面工作的最近评论, 可参看 Pearl (2000) 和 McDonald (2002).

15.2　无　向　图

一个图是指顶点和边的一个集合, $G \equiv (V, E)$. 每个顶点用一个随机向量来表示. 在本章, 随机变量有一个联合正态分布. 每个无向边是连接两个顶点的线段. 用它的两个端点来表示; 在一个无向图里, (u, v) 与 (v, u) 是相同的 (但在有向图里不是).

被一个边连接的两个顶点称为是**相邻的**; 如果不被边连接的, 则称它们是**不相邻的**. 在图 15.1 的 (a) 中, 所有顶点是不相邻的; 在 (b) 中 a 和 b 是相邻的; 在 (c) 中对子 a 和 b 以及对子 a 和 c 是相邻的; 在 (d) 中每对顶点都是相邻的.

与 G 关联的 (正态) 分布族是用加在条件分布上的一组要求 (谓以 Markov 性质) 来定义的. 因在这里考虑的分布是正态的, 所以还得有协方差阵 $\mathbf{\Sigma}$ 和它的逆 $\mathbf{\Lambda} = \mathbf{\Sigma}^{-1}$, 称之为**聚集矩阵**(concentration matrix). 然而, 许多引理和定理对非正态分布也成立. 我们将考虑三种 Markov 性的定义, 然后证明它们是等价的.

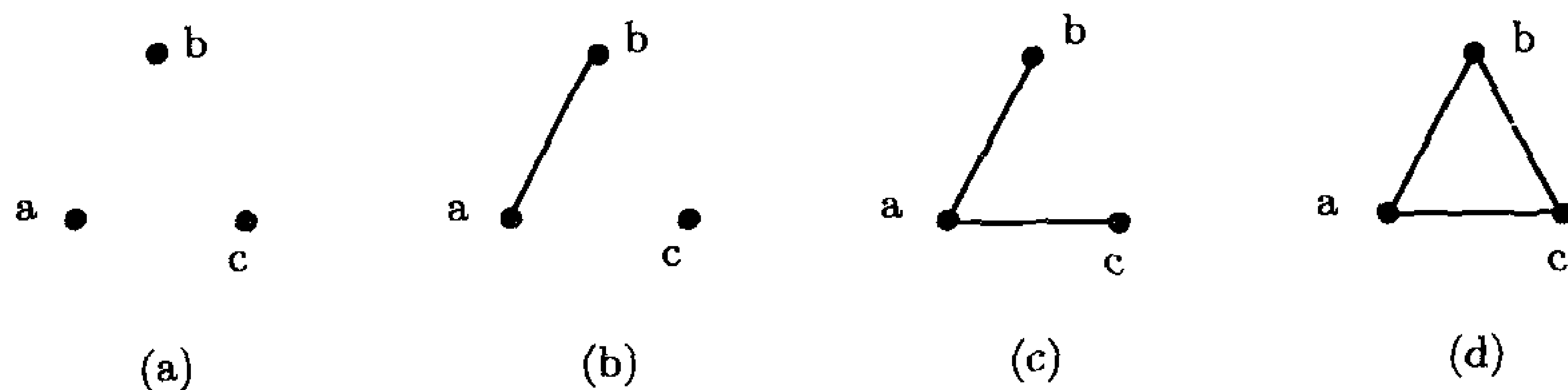


图　15.1

定义 15.2.1　*在一个图上的概率分布关于 G 是逐对马氏的, 如果对每一个不相邻的顶点对 (u, v), 在所有其他变量都在图中的条件下 $\boldsymbol{X}_u$ 和 $\boldsymbol{X}_v$ 是独立的.*

用符号表示, 即

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_v \,|\, \boldsymbol{X}_{V \setminus (u,v)}, \tag{1}$$

其中 $\amalg$ 意思指独立, $V \setminus (u, v)$ 表示 V 剔出 u 和 v 后的集合. 逐对马氏性的定义是指对于所有 $(u, v) \notin E$ 的对子 (u, v) 有 $\rho_{uv \cdot V \setminus (u,v)} = 0$. 我们也可以写为 $u \amalg v \,|\, V \setminus (u, v)$.

设 $\mathbf{\Sigma}$ 和 $\mathbf{\Lambda} = \mathbf{\Sigma}^{-1}$ 分块为

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{AA} & \mathbf{\Sigma}_{AB} \\ \mathbf{\Sigma}_{BA} & \mathbf{\Sigma}_{BB} \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{AA} & \mathbf{\Lambda}_{AB} \\ \mathbf{\Lambda}_{BA} & \mathbf{\Lambda}_{BB} \end{bmatrix}, \tag{2}$$

其中 A 和 B 是顶点的不交集合. 给定 $\boldsymbol{X}_B$ 下 $\boldsymbol{X}_A$ 的条件分布是

$$N(\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{X}_B, \boldsymbol{\Sigma}_{AA}-\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}). \tag{3}$$

条件协方差阵是

$$\boldsymbol{\Sigma}_{A\cdot B}=\boldsymbol{\Sigma}_{AA}-\boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}=\boldsymbol{\Lambda}_{AA}^{-1}. \tag{4}$$

如果 $A=(1,2)$ 和 $B=(3,\cdots,p)$, 则在给定 $X_3,\cdots,X_p$ 的条件下 X_1 和 X_2 的协方差是 $\boldsymbol{\Sigma}_{A\cdot B}=(\sigma_{ij\cdot 3\cdots p})$ 中的 $\sigma_{12\cdot 3\cdots p}$. 当且仅当 $\lambda_{12}=0$ 时它是 0, 即当且仅当 $\boldsymbol{\Lambda}_{AA}$ 是对角矩阵时 $\boldsymbol{\Sigma}_{A\cdot B}$ 是对角矩阵.

定理 15.2.1 *如果在一个图上的分布是逐对马氏的, 则对 $(i,j)\notin V$ 有 $\lambda_{ij}=0$.*

定义 15.2.2 *一个集合 A 的边界, 记为 $\mathrm{bd}(A)$, 是由那些不在 A 里但与 A 相邻的顶点组成的. A 的闭包, 记为 $\mathrm{cl}(A)$, 是 $A\cup\mathrm{bd}(A)$.*

定义 15.2.3 *一个图上的分布是局部马氏的, 如果对每个顶点 v, 变量 $\boldsymbol{X}_v$ 在给定 v 的边界条件下与不在 $\mathrm{cl}(B)$ 中的变量独立, 以记号表示为*

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_{V\backslash \mathrm{cl}(v)} \mid \boldsymbol{X}_{\mathrm{bd}(v)}. \tag{5}$$

定理 15.2.2 *条件独立性*

$$\boldsymbol{X}\amalg \boldsymbol{Y}\mid \boldsymbol{Z},\ \boldsymbol{X}\amalg \boldsymbol{Z}\mid \boldsymbol{Y} \tag{6}$$

成立的充要条件是

$$\boldsymbol{X}\amalg (\boldsymbol{Y},\boldsymbol{Z}). \tag{7}$$

证明 关系式 (6) 意味着 $\boldsymbol{X},\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 的密度可以写成为

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) &= f(\boldsymbol{x}|\boldsymbol{z})g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z}) \\ &= k(\boldsymbol{x}|\boldsymbol{y})l(\boldsymbol{z}|\boldsymbol{y})m(\boldsymbol{y}). \end{aligned} \tag{8}$$

因为 $g(\boldsymbol{y}|\boldsymbol{z})h(\boldsymbol{z})=n(\boldsymbol{y},\boldsymbol{z})=l(\boldsymbol{z}|\boldsymbol{y})m(\boldsymbol{y})$, 所以 (8) 意味着 $f(\boldsymbol{x}|\boldsymbol{z})=k(\boldsymbol{x}|\boldsymbol{y})$, 此即意味着 $f(\boldsymbol{x}|\boldsymbol{z})=k(\boldsymbol{x}|\boldsymbol{y})=p(\boldsymbol{x})$. 从而

$$f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=p(\boldsymbol{x})n(\boldsymbol{y},\boldsymbol{z}), \tag{9}$$

它就是产生 (7) 的密度. 反之, (9) 也可以写成为 (8) 的形式, 也意味着 (7). ■

推论 15.2.1 *关系式*

$$\boldsymbol{X}\amalg \boldsymbol{Y}\mid \boldsymbol{Z},\boldsymbol{W}, \qquad \boldsymbol{X}\amalg \boldsymbol{Z}\mid \boldsymbol{Y},\boldsymbol{W} \tag{10}$$

成立的充要条件是

$$\boldsymbol{X}\amalg (\boldsymbol{Y},\boldsymbol{Z})\mid \boldsymbol{W}. \tag{11}$$

定理 15.2.2 和推论 15.2.1 中的关系式有时叫做*区组独立性定理*. 它们基于正密度, 即非奇异正态分布.

定理 15.2.3 *在一个图上的局部马氏分布是逐对马氏的.*

证明 假设图是局部马氏的 (定义 15.2.3). 设 u 和 v 是非相邻顶点. 因为 v 与 u 不相邻, 所以它不在 $\mathrm{bd}(u)$ 中, 因此

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_{V\backslash \mathrm{cl}(u)} \mid \boldsymbol{X}_{\mathrm{bd}(u)}. \tag{12}$$

关系式 (12) 可以写成

$$\boldsymbol{X}_u \amalg \{\boldsymbol{X}_v, \boldsymbol{X}_{V\setminus[u,v,\mathrm{bd}(u)]}\} \mid \mathrm{bd}(u). \tag{13}$$

则推论 15.2.1 ($\boldsymbol{X} = \boldsymbol{X}_u, \boldsymbol{Y} = \boldsymbol{X}_v, \boldsymbol{Z} = \boldsymbol{Z}_{V\setminus[\mathrm{cl}(u),v]}, \boldsymbol{W} = \boldsymbol{X}_{\mathrm{bd}(u)}$) 意味着

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_v \mid \boldsymbol{X}_{V\setminus(u,v)}. \quad \blacksquare \tag{14}$$

定理 15.2.4 *在一个图上的逐对马氏分布是局部马氏的.*

证明 令 $V \setminus \mathrm{cl}(u) = v_1 \cup \cdots \cup v_n$. 则

$$u \amalg v_1 \mid \mathrm{bd}(u) \cup v_2 \cup \cdots \cup v_n, \qquad u \amalg v_2 \mid \mathrm{bd}(u) \cup v_1 \cup v_3 \cup \cdots \cup v_n, \tag{15}$$

由推论 15.2.1, 这意味着

$$u \amalg v_1 \cup v_2 \mid \mathrm{bd}(u) \cup v_3 \cup \cdots \cup v_n. \tag{16}$$

进而, 由 (16) 和

$$u \amalg v_3 \mid \mathrm{bd}(u) \cup v_1 \cup v_2 \cup v_4 \cup \cdots \cup v_n \tag{17}$$

推得

$$u \amalg v_1 \cup v_2 \cup v_3 \mid \mathrm{bd}(u) \cup v_4 \cup \cdots \cup v_n. \tag{18}$$

从而得到

$$u \amalg v_1 \cup \cdots \cup v_n \mid \mathrm{bd}(u). \tag{19}$$

第三个马氏性的概念, 即整体马氏性, 需要一些定义.

定义 15.2.4 *从 B 到 C 的一个路径是相邻顶点的一个序列 $v_0, v_1, v_2, \cdots, v_n$, 其中 $v_0 \in B$ 和 $v_n \in C$.*

定义 15.2.5 *一个集合 S 分离集合 B 和 C, 如果 S, B 和 C 是不相交的并且从 B 到 C 的每个路径都与 S 相交.*

于是, 如果对每个使得 $v_0 \in B$ 和 $v_n \in C$ 的顶点序列 $v_0, v_1, \cdots, v_n$ 至少有 $v_1, \cdots, v_{n-1}$ 中的一个是 S 中的顶点. 这里, B 和 (或)C 是非空的, 但 S 可以是空集.

定义 15.2.6 *在一个图上的分布是整体马氏的, 如果对使得 S 分离 B 和 C 的不交集合 S, B 和 C 的三元组, 在给定 $\boldsymbol{X}_S$ 条件下向量变量 $\boldsymbol{X}_B$ 和 $\boldsymbol{X}_C$ 是独立的.*

在图 15.1(c) 的例子中, a 分离 b 和 c. 如果 $\rho_{bc\cdot a} = 0$, 即 $\rho_{bc} - \rho_{ba}\rho_{ac} = 0$, 那么分布是整体马氏的. 注意到顶点的一个集合用变量的一个向量来识别的.

整体马氏性是对可能的 (正态) 分布的限制, 并且这意味着用关于它的更少的参数去作推断.

假设 $V = A \cup B \cup S$, 其中 A,B 和 S 是不相交的, 分块 $\boldsymbol{\Sigma}$ 和 $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$(聚集矩阵) 为

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{AA} & \boldsymbol{\Lambda}_{AB} & \boldsymbol{\Lambda}_{AS} \\ \boldsymbol{\Lambda}_{BA} & \boldsymbol{\Lambda}_{BB} & \boldsymbol{\Lambda}_{BS} \\ \boldsymbol{\Lambda}_{SA} & \boldsymbol{\Lambda}_{SB} & \boldsymbol{\Lambda}_{SS} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} & \boldsymbol{\Sigma}_{AS} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} & \boldsymbol{\Sigma}_{BS} \\ \boldsymbol{\Sigma}_{SA} & \boldsymbol{\Sigma}_{SB} & \boldsymbol{\Sigma}_{SS} \end{bmatrix}^{-1}. \tag{20}$$

在给定 $\boldsymbol{X}_S$ 下, $(\boldsymbol{X}'_A, \boldsymbol{X}'_B)'$ 的条件分布是正态的, 有协方差阵

$$\Sigma_{(A,B)\cdot S}=\begin{bmatrix}\Sigma_{AA} & \Sigma_{AB}\\ \Sigma_{BA} & \Sigma_{BB}\end{bmatrix}-\begin{bmatrix}\Sigma_{AS}\\ \Sigma_{BS}\end{bmatrix}\Sigma_{SS}^{-1}\begin{bmatrix}\Sigma_{SA} & \Sigma_{SB}\end{bmatrix} \tag{21}$$
$$=\begin{bmatrix}\Lambda_{AA} & \Lambda_{AB}\\ \Lambda_{BA} & \Lambda_{BB}\end{bmatrix}^{-1}.$$

定理 15.2.5 如果在一个有整体马氏分布的图上, S 分离 A 和 B, 则 $\Lambda_{AB}=0$.

证明 因为 S 分离 A 和 B, 所以 A 的每个元素 u 和 B 的每个元素 v 是不相邻的, 否则路径 (u,v) 会连接 A 和 B 而不与 S 相交. 整体马氏性是指 $\boldsymbol{X}_A$ 和 $\boldsymbol{X}_B$ 在条件分布下是不相关的, 即意味着 $\Sigma_{(A,B)\cdot S}$ 是块对角的, 因此 $\Lambda_{AB}=0$. ∎

定理 15.2.6 在一个整体马氏图上的分布是逐对马氏的.

证明 设集合 B 是 i, 集合 C 是与 i 不相邻的 j, 并且集合 A 是其余的变量. 从 B 到 C 的任何路径必须包含 A 的元素. 从而 i 和 j 在给定其他变量的条件下是独立的. ∎

定理 15.2.7 在一个图上的整体马氏分布族是局部马氏的.

证明 一个集 B 的边界分离 B 和 $V\setminus \mathrm{cl}(B)$. ∎

定理 15.2.8 在一个图上的逐对马氏分布族是整体马氏的.

证明 设 A, B 和 S 是不相交的集合, 并且依照逐对马氏图使得 S 分离 A 和 B. 令 $\#(S)$ 和 $\#(V)$ 分别记在 S 和 V 中顶点的个数. 如果 $\#(V)=\#(S)+2$, 即 $V=A\cup B\cup S$, 则必定在每个 A 和 B 中存在一个顶点, 并且逐对马氏性精确地是整体马氏性. 其余的证明是关于 $\#(S)$ 向后归纳. 假设 $\#(V)-\#(S)>2$ 并且 $V=A\cup B\cup S$. 则 A 或 B 或者两者有一个以上的顶点. 假设 A 有一个以上的顶点, 并令 $u\in A$. 则 $S\cup u$ 分离 $A\setminus u$ 和 B, 而且 $S\cup A$ 分离 u 和 B. 由归纳假设

$$\boldsymbol{X}_{A\setminus u} \amalg \boldsymbol{X}_B \mid (\boldsymbol{X}_S,\boldsymbol{X}_u),\qquad \boldsymbol{X}_u \amalg \boldsymbol{X}_B \mid (\boldsymbol{X}_S,\boldsymbol{X}_{A\setminus u}). \tag{22}$$

根据推论 15.2.1 有

$$\boldsymbol{X}_A \amalg \boldsymbol{X}_B \mid \boldsymbol{X}_S. \tag{23}$$

现在假设 $A\cup B\cup S\subset V$. 令 $u\in(V\setminus A\cup B\cup S)$. 则 $S\cup u$ 分离 A 和 B. 由归纳假设,

$$\boldsymbol{X}_A \amalg \boldsymbol{X}_B \mid (\boldsymbol{X}_S,\boldsymbol{X}_u). \tag{24}$$

另外, 要么 $A\cup S$ 分离 u 和 B 要么 $B\cup S$ 分离 A 和 u. (否则会存在一个从 B 到 u 和从 u 到 A 的路径与 S 不相交.) 如果 $A\cup S$ 分离 u 和 B, 则

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_B \mid (\boldsymbol{X}_S,\boldsymbol{X}_A). \tag{25}$$

则将推论 15.2.1 应用于 (19) 和 (20) 得到

$$(\boldsymbol{X}_A,\boldsymbol{X}_u) \amalg \boldsymbol{X}_B \mid \boldsymbol{X}_S, \tag{26}$$

由此我们推得 $\boldsymbol{X}_A \amalg \boldsymbol{X}_B \mid \boldsymbol{X}_S$. ∎

定理 15.2.3、定理 15.2.5 和定理 15.2.6 表明三种马氏性是等价的; 任何一个都可推得其他两个. 这里的证明是合理的和一般的, 但在这章假设了分布是非奇异的多元正态分布; 于是所有密度都是正的.

定义 15.2.7　一个图 $G=(V,E)$ 是完全的当且仅当在 V 中每两个顶点都是相邻的.

该定义意味着图对多元正态分布的协方差阵没有特别设定限制.

一个子集 $A\subseteq V$ 导出一个子图 $G_A=(V,E_A)$, 其中边的集合 E_A 包含 G 的所有属于 E (即满足 $(u,v)\in E$) 的边 (u,v), 其中 $u\in A$ 和 $v\in A$. 一个图的子集是完全的当且仅当 A 中每两个顶点都是在 E_A 中相邻的.

定义 15.2.8　一个团 (clique) 是一个极大完全顶点集合.

"极大" 是指如果来自 V 的另外一个顶点加进这个集合, 则这个集合就不再是完全的. 一个团可以从一个顶点 (如 v_1) 出发来构造. 如果它与任何其他顶点都不相邻, 则 v_1 独自建立一个团. 如果顶点 v_2 与 v_1 $[(v_1,v_2)\in E]$ 相邻, 则继续进行, 构造出有 v_1 和 v_2 在内的一个团, 直到找出一个极大完全子集. 于是每一个顶点是至少一个团的成员, 并且每条边包含在至少一个团内.

引理 15.2.1　如果 $\boldsymbol{X}_V$ 的分布是马氏的, 则它由所有团的边缘分布的集合所确定.

在图 15.1(a) 中 a, b, c 中的任一个都是一个团; 在 (a) 中 (a,b) 和 c 中的任一个都是一个团; 在 (c) 中 (a,b) 和 (a,c) 中的任一个都是一个团.

定义 15.2.9　称密度 $f(\boldsymbol{X}_V)$ 关于 G 是可因子分解的, 如果存在依赖于完全子图的非负函数 $g_C(\boldsymbol{X}_C)$, 使得

$$f(\boldsymbol{X}_V)=\prod_{C\text{ 完全}} g_C(\boldsymbol{X}_C). \tag{27}$$

因为只要考虑团即可, 所以一个选择性的分解是

$$f(\boldsymbol{X}_V)=\prod_{C^*\text{ 团集}} g_{C^*}(\boldsymbol{X}_{C^*}). \tag{28}$$

这些函数 $g_C(\boldsymbol{X}_C)$ 和 $g_{C^*}(\boldsymbol{X}_{C^*})$ 未必是密度或条件密度. 统计推断问题可以化为完全子图或团的问题.

定义 15.2.10　一个图的分解是由三个不相交集 A,B 和 S 形成的, 如果 $V=A\cup B\cup S$, S 分离 A 和 B, 并且 S 是完全的.

在这个定义中, 集合 A,B 和 S 中的一个或多个可能是空集. 如果 A 和 B 都是非空集, 则称该分解是真的 (proper).

定义 15.2.11　一个图是可分解的, 如果它是完全的或者存在一个真分解 (A,B,S) 把图变为可分解的子图 $G_{A\cup S}$ 和 $G_{B\cup S}$.

定理 15.2.9　假设 A,B 和 S 分解 $G=(V,E)$. 则 $\boldsymbol{X}_V$ 的密度关于 G 可分解的充要条件是它的边缘密度 $f_{A\cup S}(\boldsymbol{x}_{A\cup S})$ 和 $f_{B\cup S}(\boldsymbol{x}_{B\cup S})$ 可分解, 并且密度满足

$$f(\boldsymbol{x}_V)=\frac{f_{A\cup S}(\boldsymbol{x}_{A\cup S})f_{B\cup S}(\boldsymbol{x}_{B\cup S})}{f_S(\boldsymbol{x}_S)}. \tag{29}$$

证明 假设 $f_V(\boldsymbol{x}_V)$ 分解为

$$f_V(\boldsymbol{x}_V)=\prod_{c\in C} g_c(\boldsymbol{x}_C). \tag{30}$$

其中 $\mathscr{C}$ 是某个团集. 因为 A,B, S 分解 G, 因此每个团要么是 $A\cup S$ 的子集要么是 $B\cup S$ 的子集. 设 $\mathscr{A}$ 表示是 $A\cup S$ 的子集的团集, 而 $\mathscr{B}$ 表示是 B 的子集的团集. 则 $f_V(\boldsymbol{x}_V)=h(\boldsymbol{x}_{A\cup S})k(\boldsymbol{x}_{B\cup S})$, 其中

$$h(\boldsymbol{x}_{A\cup S})=\prod_{C\in\mathscr{A}} g_C(\boldsymbol{x}_C), \tag{31}$$

$$k(\boldsymbol{x}_{B\cup S})=\prod_{C\in\mathscr{B}\backslash\mathscr{A}} g_C(\boldsymbol{x}_C). \tag{32}$$

将 (30) 对 $C\in\mathscr{B}\backslash\mathscr{A}$ 关于 $\boldsymbol{x}_C$ 积分, 得到

$$f_{A\cup S}(\boldsymbol{x}_{A\cup S})=h(\boldsymbol{x}_{A\cup S})\bar{k}(\boldsymbol{x}_S), \tag{33}$$

其中

$$\bar{k}(\boldsymbol{x}_S)=\int k(\boldsymbol{x}_{B\cup S})\mathrm{d}\boldsymbol{x}_B. \quad ■ \tag{34}$$

进而, $f_{A\cup S}(\boldsymbol{x}_{A\cup S})$ 和 $f_{B\cup S}(\boldsymbol{x}_{B\cup S})$ 可以分解, 推得 (28).

15.3 有 向 图

我们现在考虑与方向有关系的图模型, 由一个顶点 u 表示的测量可能领先于由另一个顶点 v 表示的测量. 在图中这种*有向边*用一个从 u 到 v 的箭头来显示. 用记号 (u,v) 来表示, 它和 (v,u) 有区别. 这种优先性可以表示测量的时间, 例如, 相继两天的降雨量, 或者表示可能的因果关系.

一个考试的难度 x_1 可以影响一个学生的等级 x_3; 这种等级也是由他或她的智商 (IQ)x_2 影响的. 进而该学生的等级又影响一封推荐信的质量 x_4; IQ 是在 SAT, x_5, 上表现的一个因子. 见图 15.2.(我们将画图使得行动从左到右进行).

完全由有向边组成的图叫做*有向图*. 一个循环, 如 $1\to 2$, $2\to 3$, $3\to 1$, 很难解释, 因此通常被排除. 一个没有循环的有向图是一个*非循环有向图* (ADG 或者 DAG), 也是大家所称的非循环图. 在这章我们讨论的有向图都是非循环的.

一个非循环有向图可以表示一个*递归线性系统*. 例如, 图 15.2 可以表示

$$X_1=u_1, \tag{1}$$

$$X_2=u_2, \tag{2}$$

$$X_3=\beta_{31}X_1+\beta_{32}X_2+u_3, \tag{3}$$

$$X_4=\beta_{43}X_3+u_4, \tag{4}$$

$$X_5=\beta_{52}X_2+u_5, \tag{5}$$

其中 u_1,u_2,u_3,u_4,u_5 是相互独立的没有观测到的变量.

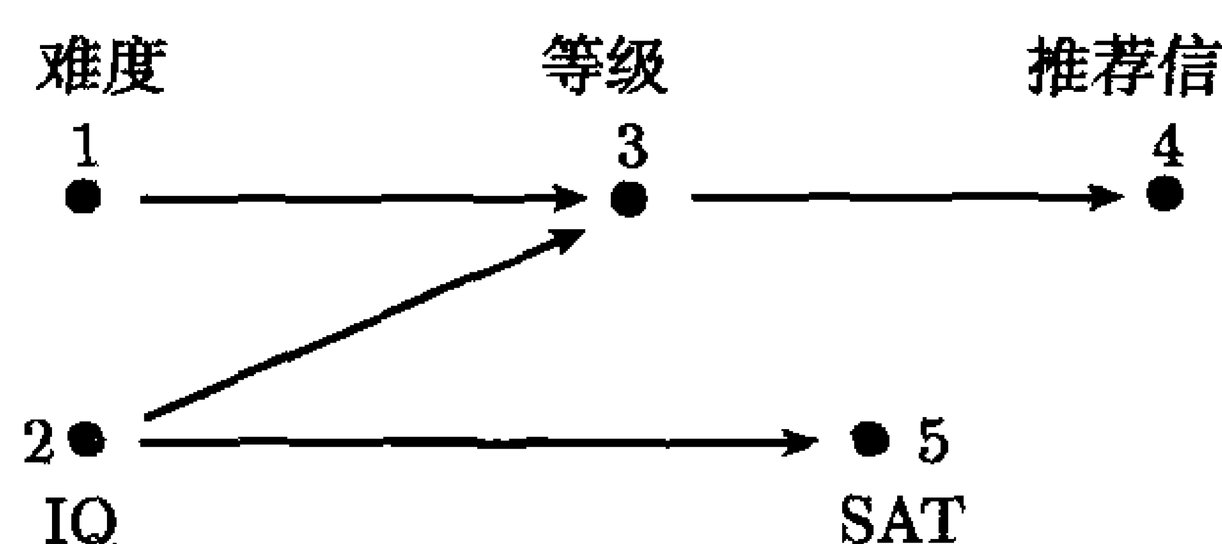


图　15.2

Wold (1960) 称这样的模型为**因果链**. 注意协方差阵是下三角的. 一般地, X_i 可以依赖于 $X_1, \cdots, X_{i-1}$.

递归线性系统 (1)~(5) 产生递归因子分解

$$f_{12345}(x_1,x_2,x_3,x_4,x_5) = f_1(x_1)f_2(x_2)f_{3|12}(x_3|x_1,x_2)f_{4|123}(x_4|x_3)f_{5|1234}(x_5|x_2). \tag{6}$$

一个有向图导出一个偏序.

定义 15.3.1　*一个非循环有向图的一个偏序 $u \leqslant v$ 是用一个有向路径来定义的:*

$$u = v_0 \to v_1 \to \cdots \to v_n = v. \tag{7}$$

这个偏序满足三条性质, 即 (i) 反身性: $v \leqslant v$; (ii) 传递性: $u \leqslant v$ 和 $v \leqslant w$ 意味着 $u \leqslant w$; (iii) 反对称性: $u \leqslant v$ 和 $v \leqslant u$ 意味着 $u = w$. 进而, $u \leqslant v$ 和 $u \neq v$ 意味着 $u < v$.

定义 15.3.2　*如果 $u \to v$, 则 u 是 v 的父辈, 称 $u = \mathrm{pa}(v)$, 而 v 是 u 的子辈, 称 $v = \mathrm{ch}(u)$. 以符号表示*

$$\mathrm{pa}(v) = \{w \in V \setminus v | w \to v\}, \tag{8}$$

$$\mathrm{ch}(u) = \{w \in V \setminus u | u \to w\}. \tag{9}$$

在图 15.2 展示的图中, 我们有 $(1,2) = \mathrm{pa}(3)$, $3 = \mathrm{pa}(4)$, $2 = \mathrm{pa}(5)$, $3 = \mathrm{ch}(1,2)$, $4 = \mathrm{ch}(3)$, $5 = \mathrm{ch}(2)$.

定义 15.3.3　*如果 $u < v$, 则 v 是 u 的后代,*

$$\mathrm{de}(u) = \{v | u < v\}, \tag{10}$$

u 是 v 的祖先,

$$\mathrm{an}(v) = \{u | u < v\}. \tag{11}$$

u 的非后代集是 $\mathrm{Nd}(u) = V \setminus \mathrm{de}(u)$, 严格非后代集是 $\mathrm{Nd}(u) = \mathrm{Nd}(u) \setminus u$. 定义 $\mathrm{An}(A) = \mathrm{an}(A) \cup A$.

注意

$$\mathrm{pa}(v) \subseteq \mathrm{an}(v) \subseteq \mathrm{nd}(v). \tag{12}$$

在无向图研究中, 我们考虑独立定义的三种马氏性, 然后证明了有一种马氏性的图也有其他两种马氏性. 在非循环有向图情形, 我们将定义三种类似的马氏性, 但定义是不同的, 因为它们考虑到行动的方向或者影响.

定义 15.3.4　*在一个非循环有向图 G 上的分布是逐对马氏的, 如果对每个 $v \in V$ 和 $w \in \mathrm{nd}(v) \setminus \mathrm{pa}(v)$,*

$$v \perp\!\!\!\perp w \mid \mathrm{nd}(v) \setminus w. \tag{13}$$

与无向图情形的定义 15.2.1 相比, 注意到, 注意力仅放在 $\mathrm{nd}(v)$ 的顶点上; 因为 $\mathrm{pa}(v)$ 是 v 的有效边界, 顶点 w 和 v 是不相邻的. (参看图 15.3.) 还注意条件集包含 v 的父辈, 但不包含 v 的子辈 (是非后代).

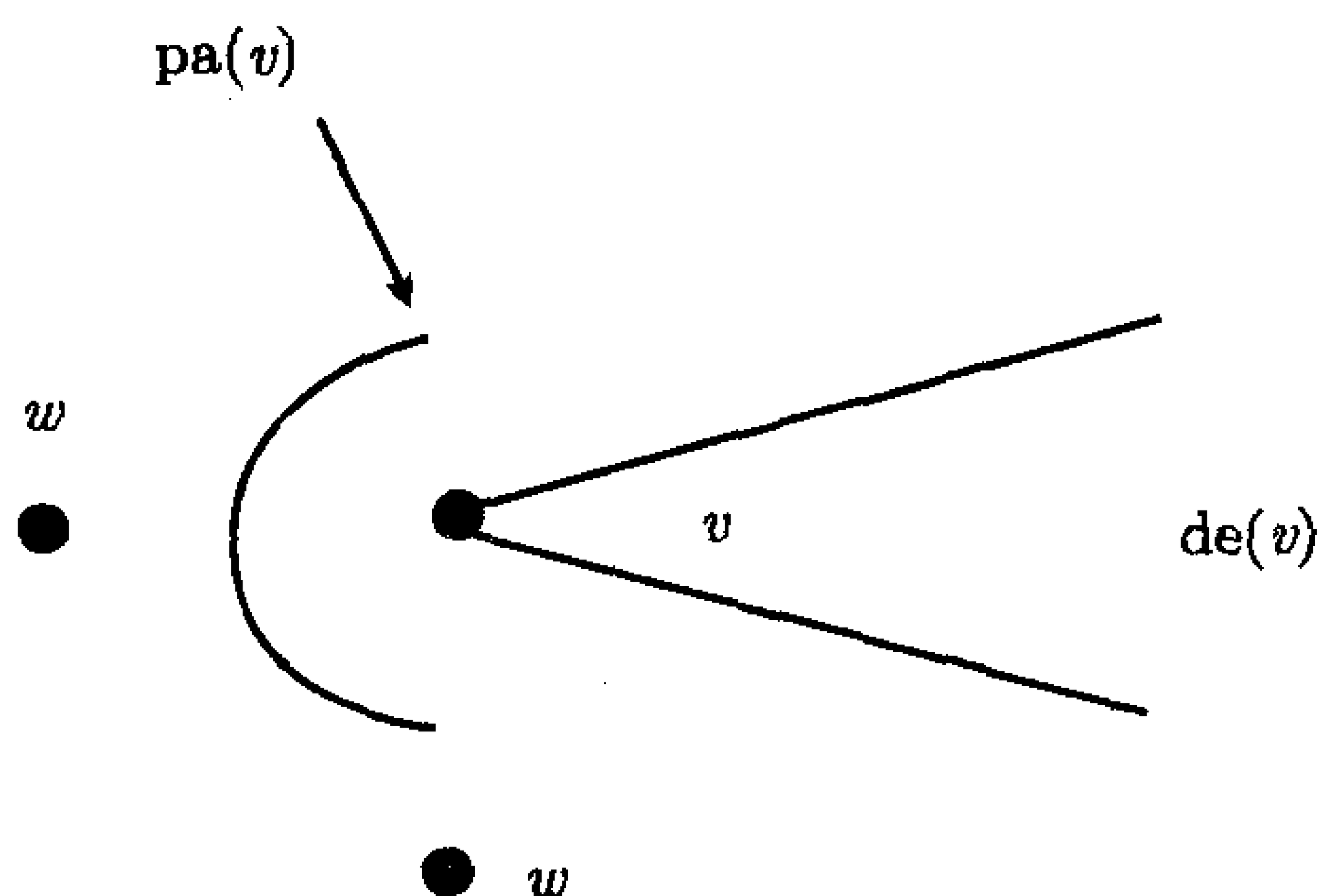


图 15.3

定义 15.3.5 *在一个非循环有向图 G 上的分布是局部马氏的, 如果*

$$v \perp\!\!\!\perp [\mathrm{nd}(v) \setminus \mathrm{pa}(v)] \mid \mathrm{pa}(v). \tag{14}$$

在局部马氏性的定义中, 条件仅仅加在 v 的父辈上, 而在逐对马氏性的定义中, 条件是加在所有其他非后代上的. 这些属性对应着无向图情形的定义 15.2.1 和定义 15.2.3.

在图 15.2 中, 我们有 $1 \perp\!\!\!\perp 2,5$, $3 \perp\!\!\!\perp 5 \mid 2$, $4 \perp\!\!\!\perp 1,2,5 \mid 3$, $5 \perp\!\!\!\perp 1,3,4 \mid 2$. 在图 15.2 中用线代替箭头 (用无向边代替有向边), 从而构造出一个无向图, 在这图上的一个局部马氏分布应该含有条件独立性 $1 \perp\!\!\!\perp 2 \mid 3$, $1,2 \perp\!\!\!\perp 4 \mid 3$, $1,3,4 \perp\!\!\!\perp 5$. 在箭头的表示中, 表示时间序列 X_4 涉及 (X_2, X_3) 的将来; 对将来不能加条件.

作为另一个例子, 考虑一个自回归时间序列 $y_0, y_1, \cdots, y_T$, 其定义为

$$y_i = \rho y_{i-1} + u_i, \quad i = 1, 2, \cdots, T, \tag{15}$$

其中 $u_1, \cdots, u_T$ 是独立的 $N(0, \sigma^2)$ 变量, y_0 有分布 $N[0, \sigma^2/(1+\rho)^2]$. 在给定 y_t 下, 将来 $y_{i+1}, \cdots, y_T$ 与过去 $y_0, \cdots, y_{i-1}$ 独立.

定理 15.3.1 *在一个非循环有向图上的局部马氏分布是逐对马氏的.*

证明 该证明与无向图情形的定理 15.2.3 的证明相同. ■

定理 15.3.2 *在一个非循环有向图上的逐对马氏分布是局部马氏的.*

证明 该证明与无向图情形的定理 15.2.4 的证明相同. ■

另一个马氏性基于顺序地把在一个反映行动的方向或者导致的偏序上的顶点编号.

定义 15.3.6 *V 的元素的一个列举* (enumeration) *是完好编号* (well-numbered) *的, 如果 $i < j \Rightarrow v_j \nless v_i$, 或者等价地 $v_j < v_i \Rightarrow j < i$.*

定理 15.3.3 一个有限有序集 $(V, \leqslant)$ 容许至少一个完好编号.

定义 15.3.7 一个元素 $a^* \in V$ 是极大的 (或终点的), 如果 $a^* \leqslant b \Rightarrow a^* = b$.

引理 15.3.1 一个有限偏序集 $(V, \leqslant)$ 至少有一个极大元素 a^*.

引理的证明 用归纳法证明. 如果 $\#(V) = 1$, 那么 $a^* = a$. 现假设该引理对 $\#(V) = n$ 成立, 考虑 $\#(V) = n + 1$. 则对任意 $a \in V$ 有 $V = a \cup (V \setminus a)$. 因为 $\#(V \setminus a) = n$, 所以 $V \setminus a$ 是极大元素, 设为 $\bar{a}$. 那么, 要么 $\bar{a} \leqslant a$ 并因此 a 是极大的, 要么 $\bar{a} \neq a$ 并因此 $\bar{a}$ 是极大的. ■

定理 15.3.3 的证明 我们将构造一个完好编号. 设 v^* 是一个极大元素; 定义 $v_n = v^*$. 在 $V \setminus v_n$ 中设 v^{**} 是一个极大元素; 定义 $v_{n-1} = v^{**}$. 在第 j 步设 v^{***} 是在 $V \setminus (v_n, \cdots, v_{n-j+1})$ 中的一个极大元素; 定义 $v_{n-j} = v^{***}, j = 3, \cdots, n-1$. 则 $v_1 = V \setminus (v_1, \cdots, v_{n-1})$. 这个构造满足定义 15.3.6. ■

作为 V 的完好编号 $v^{(1)}, \cdots, v^{(n)}$ 意味着, 在任意一个有向路径 $u = v^{(i_0)} \to v^{(i_1)} \to \cdots \to v^{(i_n)} = v$ 中这些标号满足 $i_0 \leqslant i_1 \leqslant \cdots i_n$. 完好编号不必是唯一的. 因为 V 是有限的, 所以一个极大元素总可以通过比较 v_i 和 v_j 找到, 因为至多 $n(n-1)/2$ 个对.

定义 15.3.8 设 $\{v_1, \cdots, v_n\}$ 是一个非循环有向图 G 上的一个完好编号. 在 G 上的一个分布是关于这个完好编号马氏的, 如果

$$v_i \amalg (v_1, \cdots, v_{i-1}) \setminus \mathrm{pa}(v_i) \,|\, \mathrm{pa}(v_i), \quad i = 3, \cdots, n. \tag{16}$$

表面上, 该定义依赖于完好编号的选择, 但是由定理 15.3.4 知, 并不是这样.

定理 15.3.4 一个完好编号的马氏非循环有向图上的一个分布是局部马氏的.

证明 $(v_1, \cdots, v_{i-1}) \in \mathrm{nd}(v_i) \setminus \mathrm{pa}(v_i)$. ■

定义 15.3.9 一个非循环有向图 $G = (V, E)$ 的理念图 (moral graph)G^m 是由在每个顶点 $v \in V$ 的父辈之间加 (无向) 边和用一个无向边代替每个有向边来构造的无向图.

在图论的行话里, 一个顶点的父辈 (parents) 是“已婚的”.

定义 15.3.10 在一个非循环有向图上的分布是整体马氏的, 如果对使得在 $[G_{\mathrm{An}(A \cup B \cup S)}]^m$ 中 S 分离 A 和 B 的每个三元组 A, B 和 S, 有 $A \amalg B | S$.

定理 15.3.5 在一个为整体马氏的非循环有向图上的分布是局部马氏的.

证明 对任意 $v \in V$, 设在整体马氏性的定义中 $\mathrm{pa}(v) = S$. 又设 $v = A$ 和 $\mathrm{nd}(v) \setminus \mathrm{pa}(v) = B$. 一个顶点 $w \in \mathrm{nd}(v) \setminus \mathrm{pa}(v)$ 是 $\mathrm{An}(A \cup B \cup S)$ 中的一个顶点. 令 $\pi = w = v_0, v_1 \cdots, v_n = v$ 是 $[G_{\mathrm{An}(A \cup B \cup S)}]^m = [G_{\mathrm{Nd}(v)}]^m$ 中一个从 w 到 v 的路径. 如果 (v_{n-1}, v_n) 对应 $G_{\mathrm{Nd}(v)}]^m$ 中的一个有向边 $(v_{n-1} \to v_n)$, 则 $v_{n-1} \in \mathrm{pa}(v) = S$ 并且 $\mathrm{pa}(v)$ 分离 $\mathrm{nd}(v) \setminus \mathrm{pa}(v)$ 和 v. [该有向边 $(v_{n-1} \leftarrow v_n)$ 意味着 $v_{n-1} \in \mathrm{de}(v)$.] ■

定理 15.3.6 在一个为局部马氏的非循环有向图上的分布是整体马氏的.

证明很长, 在此省略.

递归因子分解

非循环有向图的递归方面允许密度的一个系统分解. 利用定理 15.3.4 的构造. 设 $n=|V|$, 则 v_n 是 V 的一个极大元素. 则

$$\boldsymbol{X}_{V\backslash \mathrm{cl}(v_n)} \amalg \boldsymbol{X}_{v_n} \,|\, \mathrm{pa}(v_n). \tag{17}$$

于是 (在正态情况下)

$$\mathrm{E}(\boldsymbol{X}_{v_n} \,|\, \mathrm{pa}(v_n)) = \boldsymbol{\alpha}_n + \boldsymbol{B}_n X_{\mathrm{pa}(v_n)}. \tag{18}$$

$$\mathrm{E}(\boldsymbol{X}_{v_n} - \mathrm{E}(\boldsymbol{X}_{v_n}))(\boldsymbol{X}_{v_n} - \mathrm{E}(\boldsymbol{X}_{v_n}))' = \boldsymbol{\Sigma}_n. \tag{19}$$

在第 j 步设 v_{n-j+1} 是 $V\setminus(v_n,\cdots,v_{n-j+2})$ 的极大元素. 则

$$\boldsymbol{X}_{V\backslash[v_n,\cdots,v_{n-j+2},\mathrm{cl}(v_{n-j+1})]} \amalg \boldsymbol{X}_{v_{n-j+1}} \,|\, \mathrm{pa}(v_{n-j+1}). \tag{20}$$

于是

$$\mathrm{E}(\boldsymbol{X}_{v_n-j+1} \,|\, \mathrm{pa}(v_{n-j+1})) = \boldsymbol{\alpha}_{n-j+1} + \boldsymbol{B}_{n-j+1}\boldsymbol{X}_{\mathrm{pa}(v_{n-j+1})}. \tag{21}$$

$$\mathrm{E}(\boldsymbol{X}_{v_{n-j+1}} - \mathrm{E}(\boldsymbol{X}_{v_{n-j+1}}))(\boldsymbol{X}_{v_{n-j+1}} - \mathrm{E}(\boldsymbol{X}_{v_{n-j+1}}))' = \boldsymbol{\Sigma}_{n-j+1},\ j=1,\cdots,n-1. \tag{22}$$

向量 $\boldsymbol{X}_{v_{n-j+1}}$ 与 $\mathrm{pa}(v_{n-j+1})$ 独立. 从 (18) 到 (22) 的关系可以写成为生成方程. 设

$$\boldsymbol{x}_1 = \boldsymbol{\alpha}_1 + \boldsymbol{\varepsilon}_1, \tag{23}$$

$$\boldsymbol{x}_2 = \boldsymbol{\alpha}_2 + \boldsymbol{B}_2\boldsymbol{x}_1 + \boldsymbol{\varepsilon}_2, \tag{24}$$

$$\vdots$$

$$\boldsymbol{x}_{n-1} = \boldsymbol{\alpha}_{n-1} + \boldsymbol{B}_{n-1}(\boldsymbol{x}_1',\cdots,\boldsymbol{x}_{n-2}')' + \boldsymbol{\varepsilon}_{n-1}, \tag{25}$$

$$\boldsymbol{x}_n = \boldsymbol{\alpha}_n + \boldsymbol{B}_n(\boldsymbol{x}_1',\cdots,\boldsymbol{x}_{n-1}')' + \boldsymbol{\varepsilon}_n, \tag{26}$$

其中 $\varepsilon_1,\cdots,\varepsilon_n$ 是独立的随机向量, 有 $\mathrm{E}(\varepsilon_j\varepsilon_j')=\boldsymbol{\Sigma}_j$. 从 (23) 到 (26) 的矩阵是

$$\boldsymbol{B}\boldsymbol{x} = \boldsymbol{\alpha} + \boldsymbol{\varepsilon}. \tag{27}$$

其中

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \vdots \\ \boldsymbol{\alpha}_n \end{bmatrix},\quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{I} & 0 & 0 & \cdots & 0 \\ -\boldsymbol{B}_{21} & \boldsymbol{I} & 0 & \cdots & 0 \\ -\boldsymbol{B}_{31} & -\boldsymbol{B}_{32} & \boldsymbol{I} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -\boldsymbol{B}_{n1} & -\boldsymbol{B}_{n2} & -\boldsymbol{B}_{n3} & \cdots & \boldsymbol{I} \end{bmatrix},\quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \tag{28}$$

并且 $\boldsymbol{B}_{ji}=0$ 如果 $i<j-k_j$. 因为 $\boldsymbol{B}$ 的行列式是 1. (27) 可以解出, 有

$$\boldsymbol{x} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Gamma}^{-1}\boldsymbol{\varepsilon}. \tag{29}$$

矩阵 $\boldsymbol{\Gamma}^{-1}$ 也是下三角的.

15.4 链图

一个链图包含有无向和有向两种边, 但仅仅一定的顶点和边的模式是允许的. 假设图 $G=(V,E)$ 的顶点集 V 可以被划分成子集 $V=V(1)\cup\cdots\cup V(T)$, 使得在一个子集内顶点由无向边连接, 而有向边连接在不同子集的顶点. 设 $\mathscr{T}(G)$ 是顶点 $1,\cdots,T$ 的集合, 又设 $\mathscr{E}(G)$ 是 (有向边) 的集合, 使得 $\tau\to\sigma$ 当且仅当至少存在一个元素 $u\in V(\tau)$ 和至少存在一个元素 $u\in V(\sigma)$ 使得 $u\to v$ 是在 $E(G$ 的边集) 中. 则 $\mathscr{D}(G)=[\mathscr{T}(G),\mathscr{E}(G)]$ 是一个非循环有向图, 我们可以对 $\mathscr{D}(G)$ 定义 $\mathrm{pa}_{\mathscr{D}}(\tau)$, 等等.

设 $\boldsymbol{X}_\tau=\{\boldsymbol{X}_u\,|\,u\in V(\tau)\}$. 在一个集合内, 顶点形成与给定过去 (即较早的集合) 下的概率分布有关的一个无向图. 参见图 15.4 [Lauritzen (1996)] 和图 15.5.

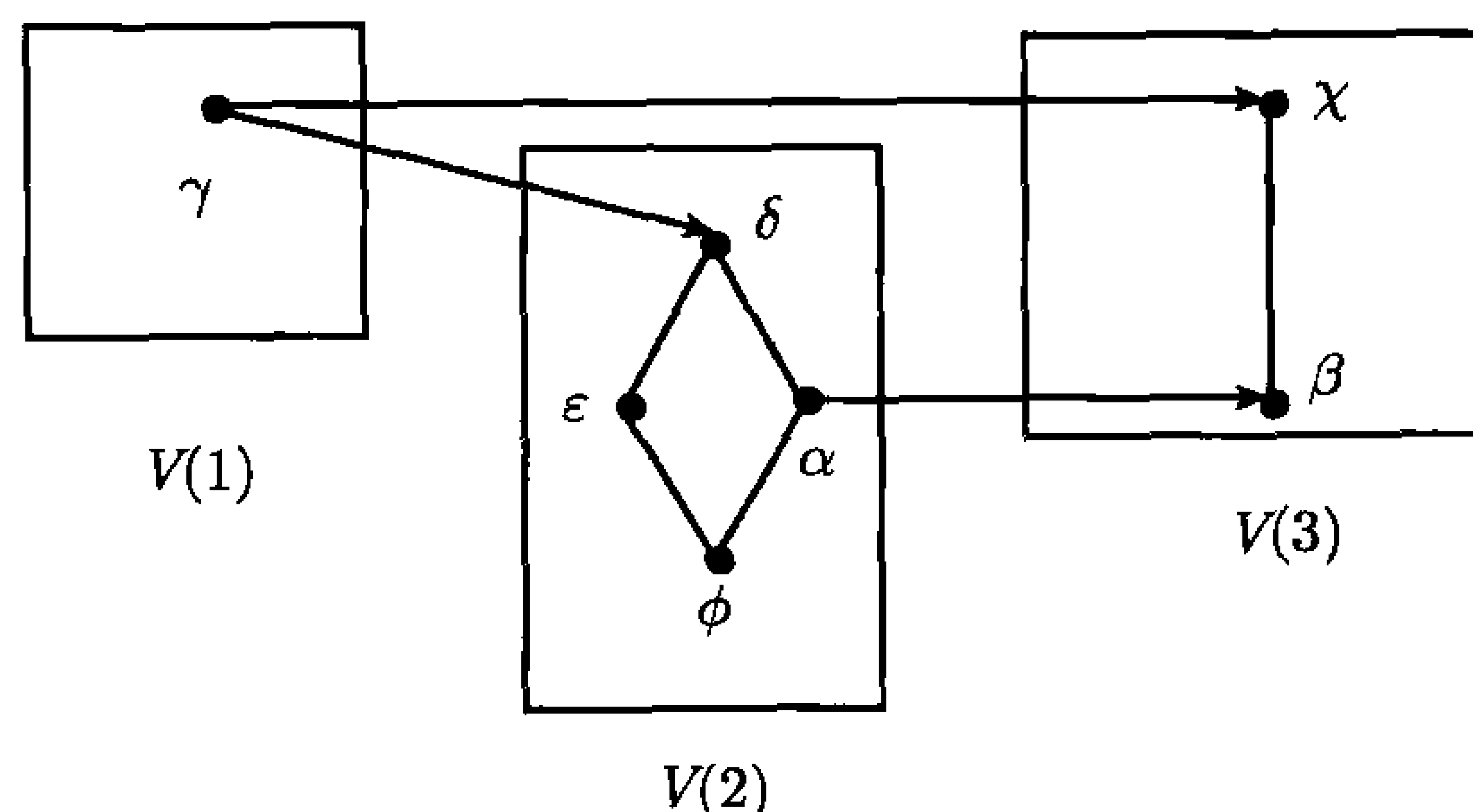


图 15.4 一个链图

我们现在定义由 Lauritzen and Wermuth (1989) 和 Frydenberg (1990) 确定的马氏性.

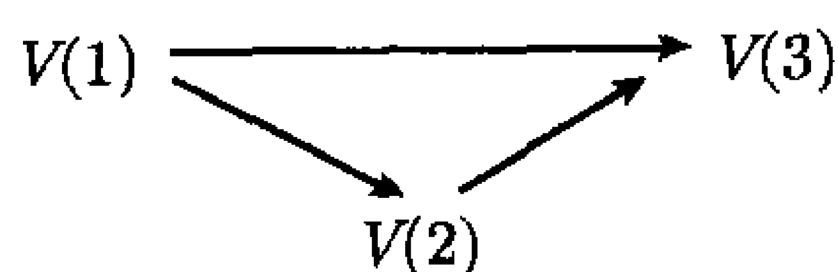


图 15.5 在 $V=V(1)\cup V(2)\cup V(3)$ 上对应的诱导非循环有向图

(C1) $\boldsymbol{X}_\tau(\tau=1,\cdots,T)$ 的分布是关于非循环有向图 $\mathscr{D}(G)$ 局部马氏的, 即

$$\boldsymbol{X}_\tau \amalg \boldsymbol{X}_\sigma\,|\,\boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)},\quad \sigma\in\mathrm{nd}_{\mathscr{D}}(\tau)\backslash\mathrm{pa}_{\mathscr{D}}(\tau).\quad(1)$$

(C2) 对每个 τ, $\boldsymbol{X}_\tau$ 在给定 $\boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$ 下的条件分布是关于 $V(\tau)$ 上无向图整体马氏的.

(C3)
$$\boldsymbol{X}_u \amalg \boldsymbol{X}_v\,|\,\boldsymbol{X}_{\mathrm{bd}_G}(U),\ \ u\in U\subseteq V(\tau),\ \ v\in\mathrm{pa}_{\mathscr{D}}(\tau)\setminus\mathrm{pa}_G(U).\qquad(2)$$

这里 $\mathrm{bd}_G(U)=\mathrm{pa}_G(U)\cup\mathrm{nb}_G(U)$. 在满足 (C1), (C2) 和 (C3) 的链图 G 上的一个分布是 LWF 区组递归马氏的.

在图 15.6 中, 有 $\mathrm{pa}_{\mathscr{D}}(\tau)=\{\tau-1,\tau-2\}$ 和 $\mathrm{nd}_{\mathscr{D}}(\tau)\backslash\mathrm{pa}_{\mathscr{D}}(\tau)=\{\tau-3,\tau-4,\cdots,1\}$. 集合 $U=\{u,w\}$ 是 $V(\tau)$ 中的一个集合, 而 $\mathrm{pa}_G(U)$ 是 $V(\tau-1)\cup V(\tau-2)$ 中的一

个集合, 对 $u \in U$, 它包含 $\mathrm{pa}_G(u)$, 即 $\mathrm{pa}_G(U) = \{x, y\}$.

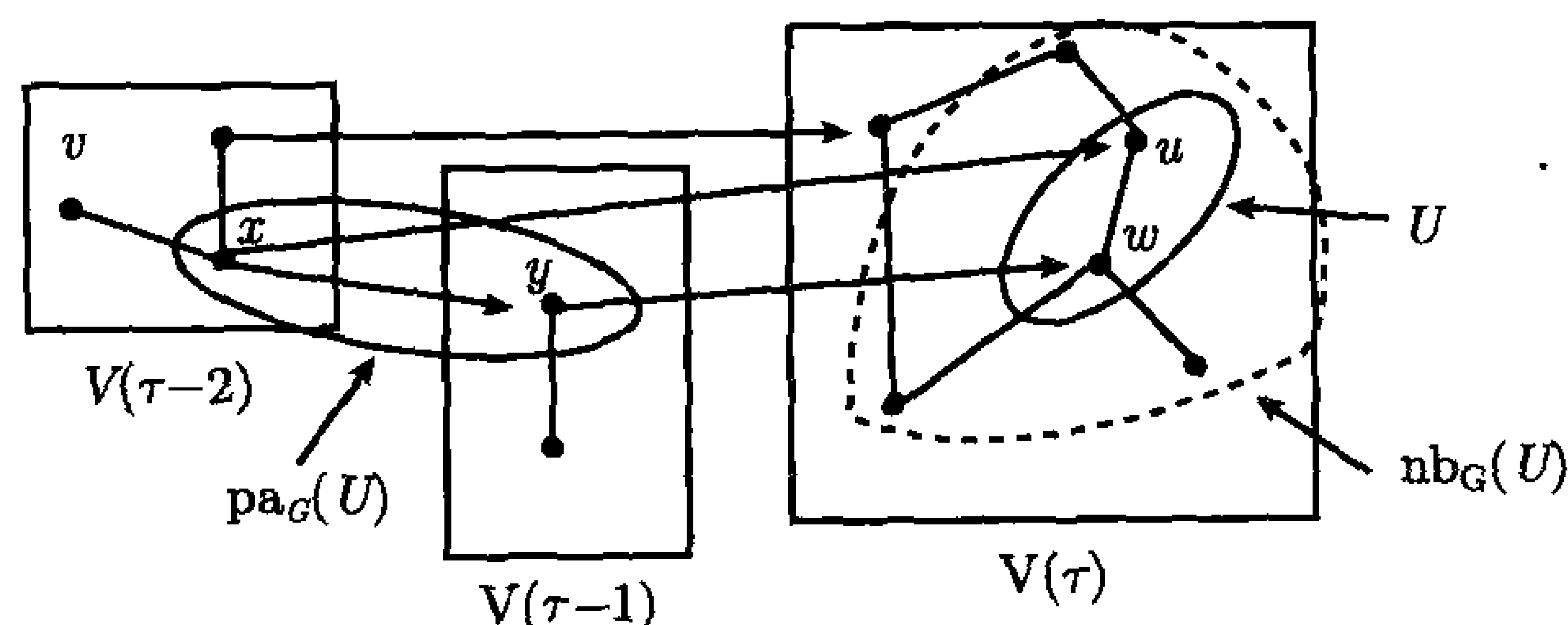


图 15.6　一个链图

Andersson, Madigan, and Perlman (2001) 提出另一种马氏性 (AMP), (C3) 由下式代替,

(C3*)

$$\boldsymbol{X}_u \amalg \boldsymbol{X}_v \mid \boldsymbol{X}_{\mathrm{pa}_G(U)}, \quad u \in U \subseteq V(\tau), \quad v \in \mathrm{pa}_{\mathscr{D}}(\tau) \setminus \mathrm{pa}_G(U). \tag{3}$$

在图 15.6 中, 对于在 $V(\tau-2) \cup V(\tau-1)$ 中的一个顶点 v, 当在 $\boldsymbol{X}_{\mathrm{pa}_G(U)} = (\boldsymbol{X}_x, \boldsymbol{X}_y)$ 上回归时, $\boldsymbol{X}_v$ 是与 $\boldsymbol{X}_u$ $[u \in U \subseteq V(\tau)]$ 条件独立的. (C3) 和 (C3*) 的差别是, 在 (C3) 中的条件是加在 $\mathrm{bd}_G(U) = \mathrm{pa}_G(U) \cup \mathrm{nb}_G(U)$ 上的, 而在 (C3*) 中的条件是仅仅加在 $\mathrm{pa}_G(U)$ 上的. 参见图 15.6. 在 (C3*) 中的条件是加在过去的变量上. 图 15.7 [Andersson, Madigan, and Perlman (2001)] 解释了 LWF 马氏性和 AMP 马氏性之间的差别,

$$\text{LWF}: \quad \boldsymbol{X}_1 \amalg \boldsymbol{X}_4 \mid \boldsymbol{X}_2, \boldsymbol{X}_3, \quad \boldsymbol{X}_2 \amalg \boldsymbol{X}_3 \mid \boldsymbol{X}_1, \boldsymbol{X}_4, \tag{4}$$

$$\text{AMP}: \quad \boldsymbol{X}_1 \amalg \boldsymbol{X}_4 \mid \boldsymbol{X}_2, \quad \boldsymbol{X}_2 \amalg \boldsymbol{X}_3 \mid \boldsymbol{X}_1. \tag{5}$$

注意在 (5) 中 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_4$ 是在给定 $\boldsymbol{X}_2$ 下条件独立的, $\boldsymbol{X}_4$ 的条件分布依赖于 $\mathrm{pa}(v_2)$, 而不依赖于 $\boldsymbol{X}_3$.

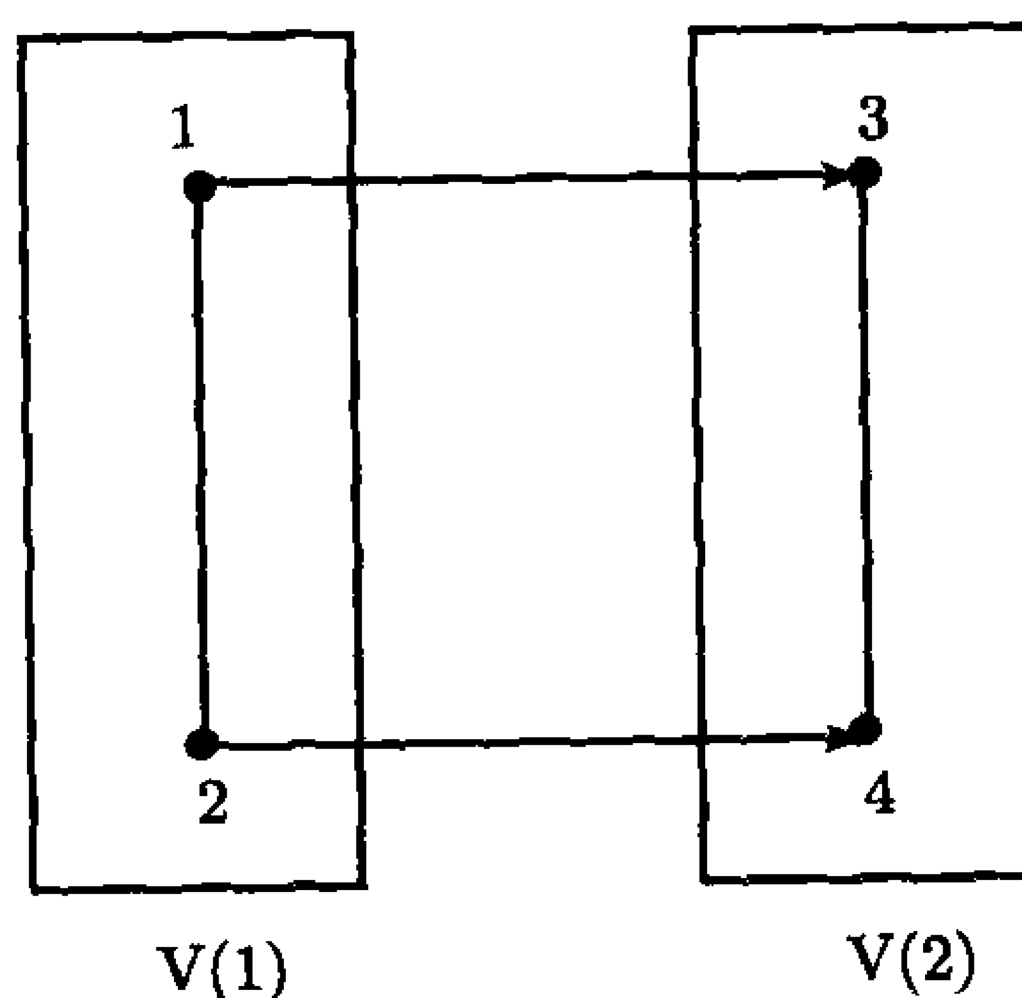


图 15.7　一个链图

AMP 指定的马氏性允许用一个区组递归方程来阐述. 在图 15.7 的例子中, 标量 X_1 和 X_2 $[v_1, v_2 \in V(1)]$ 的分布可以特定为

$$X_1 = \varepsilon_1, \tag{6}$$

$$X_2 = \varepsilon_2, \tag{7}$$

其中 $(\varepsilon_1, \varepsilon_2)$ 有任意的 (正态) 分布. 因为 X_3 直接依赖于 X_1 而 X_4 直接依赖于 X_2, 所以我们可以写

$$X_3 = \beta_{31} X_1 + \varepsilon_3, \tag{8}$$

$$X_4 = \beta_{42} X_2 + \varepsilon_4, \tag{9}$$

其中 $(\varepsilon_3, \varepsilon_4)$ 有与 $(\varepsilon_1, \varepsilon_2)$ 独立的任意分布, 并因此与 (X_1, X_2) 独立.

一般地, AMP 模型可以表达为 15.3 节的 (26) 式.

15.5　统 计 推 断

15.5.1　正态分布

设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ 是在具有分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 的 $\boldsymbol{X}$ 上的 N 个观测. 记 $\bar{\boldsymbol{x}} = N^{-1}\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha$ 和 $\boldsymbol{S} = (N-1)^{-1}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}})' = (N-1)^{-1}[\sum_{\alpha=1}^{N} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha' - N\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}']$. 似然函数是

$$\begin{aligned}&(2\pi)^{-Np/2}|\boldsymbol{\Sigma}|^{-N/2}\exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_\alpha - \boldsymbol{\mu})\right\} \\ =&(2\pi)^{-Np/2}|\boldsymbol{\Sigma}|^{-N/2}\exp\left\{-\frac{1}{2}[(N-1)\mathrm{tr}\boldsymbol{S}\boldsymbol{\Sigma}^{-1} + N(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})]\right\}.\end{aligned} \tag{1}$$

上面的形式表明 $\bar{\boldsymbol{x}}$ 和 $\boldsymbol{S}$ 是 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ 的一对充分统计量, 并且他们的分布是独立的. 本章的关注点是在独立性上, 其仅依赖于协方差阵 $\boldsymbol{\Sigma}$ 而不依赖于 $\boldsymbol{\mu}$. 对本章的其余部分, 我们将注意力放在均值上. 因此, 我们假设总体分布是 $N(\boldsymbol{0}, \boldsymbol{\Sigma})$, 样本是 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, 以及 $\boldsymbol{S} = (1/n)\sum_{\alpha=1}^{n} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha'$. 似然函数可以写成

$$(2\pi)^{-np/2}|\boldsymbol{\Lambda}|^{pn/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\, n\boldsymbol{\Lambda}\boldsymbol{S}\right\} = \exp\left[-\Psi(\boldsymbol{\Lambda}) - \frac{1}{2}\sum_{i=1}^{p}\lambda_{ii}t_{ii} - \sum_{i<j}\lambda_{ij}t_{ij}\right], \tag{2}$$

其中 $\boldsymbol{\Lambda} = (\lambda_{ij}) = \boldsymbol{\Sigma}^{-1}, \boldsymbol{T} = (t_{ij}) = \sum_{\alpha=1}^{n} \boldsymbol{x}_\alpha \boldsymbol{x}_\alpha'$, 并且 $\Psi(\boldsymbol{\Lambda}) = \frac{1}{2}pn\ln(2\pi) - \frac{1}{2}n\ln|\boldsymbol{\Lambda}|$.

该似然是在带有典范参数 $\boldsymbol{\Lambda}$ 和统计量 $\boldsymbol{T}$ 的指数分布族中. 在没有限制时, $\boldsymbol{\Sigma}$ 的极大似然估计是 $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S} = (1/n)\boldsymbol{T}$. 因为 $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ 是 $\boldsymbol{\Sigma}$ 的一对一变换, 所以 $\boldsymbol{\Lambda}$ 的极大似然估计是 $\hat{\boldsymbol{\Lambda}} = \hat{\boldsymbol{\Sigma}}^{-1}$.

15.5.2　协方差选择模型

在无向图中, 许多模型涉及在 $\boldsymbol{\Lambda}$ 的元素上要求为零的限制. Dempster (1972) 研究了这样的模型并引进了术语*协方差选择*. 当 (有向) 图满足逐对马氏条件时, 对 $(i,j) \notin E$ 有 $\lambda_{ij} = 0$. 这里假设图满足马氏. 进而我们假设 $n \geqslant p$, 则 S 以概率 1 是正定的.

似然函数是

$$(2\pi)^{-p/2}|\boldsymbol{\Lambda}|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\lambda_{ii}+\sum_{(i,j)\in E}\lambda_{ij}ns_{ij}\right\},\tag{3}$$

其中 $\boldsymbol{\Lambda}$ 满足条件 $\lambda_{ij}=0,(i,j)\notin E$. 在这个形式中**典范参数**是 $\lambda_{11},\cdots,\lambda_{pp}$ 和 $\lambda_{ij},(i,j)\in E$. **典范变量**是 $s_{11},\cdots,s_{pp}$ 和 $s_{ij},(i,j)\in E$; 这些形成统计量的一个充分集. 为极大化似然函数, 我们对 (3) 关于 $\lambda_{ii}(i=1,\cdots,p)$ 和 $\lambda_{ij}((i,j)\in E)$ 求导数, 便得到方程 (4) 和 (5).

定理 15.5.1 *在模型 (3) 中 $\boldsymbol{\Sigma}$ 的极大似然估计是由以下两式给出:*

$$\hat{\sigma}_{ij}=s_{ij},\quad i=j \text{ 或者 } (i,j)\in E,\tag{4}$$

$$\lambda_{ij}=0,\quad i\neq j \text{ 并且 } (i,j)\notin E,\tag{5}$$

其中 $\boldsymbol{\Lambda}=\boldsymbol{\Sigma}^{-1}$.

这个可由指数族的一般理论得到. 参见 Lauritzen (1996), 定理 5.3 和附录 D.1.

这里我们将证明, 对一个可分解的图, 方程 (4) 和 (5) 通过对计算发展一种算法有唯一正定解. 我们按照 Speed and Kiiveri (1986) 的证明.

定理 15.5.2 *设 $\boldsymbol{L}$ 和 $\boldsymbol{M}$ 是 $p\times p$ 的正定矩阵. 存在一个唯一的正定矩阵 $\boldsymbol{K}$ 使得*

$$k_{ij}=l_{ij},\qquad i=j \text{ 或者 } (i,j)\in E,\tag{6}$$

$$k^{ij}=m^{ij},\qquad i\neq j \text{ 并且 } (i,j)\notin E,\tag{7}$$

其中 $(k^{ij})=\boldsymbol{K}^{-1}$ 和 $(m^{ij})=\boldsymbol{M}^{-1}$.

定理 15.5.2 的证明依赖于几个引理. 在极大似然估计中 $\boldsymbol{L}=\boldsymbol{S}$, $\boldsymbol{M}=\boldsymbol{I}$ 或者任何其他的对角矩阵, 以及 $\boldsymbol{K}=\hat{\boldsymbol{\Sigma}}$. 为拓展这个方面, 我们利用 Kullback 信息. 对一对多元正态分布 $N(\mathbf{0},\boldsymbol{P})$ 和 $N(\mathbf{0},\boldsymbol{R})$ 定义

$$\boldsymbol{I}(\boldsymbol{P}\,|\,\boldsymbol{R})=\mathrm{E}_p\ln\frac{n(\boldsymbol{x}\,|\,\mathbf{0},\boldsymbol{P})}{n(\boldsymbol{x}\,|\,\mathbf{0},\boldsymbol{R})}=-\frac{1}{2}\left[\ln|\boldsymbol{P}\boldsymbol{R}^{-1}|+\mathrm{tr}\,(\boldsymbol{I}-\boldsymbol{P}\boldsymbol{R}^{-1})\right].\tag{8}$$

引理 15.5.1 *假设 $\boldsymbol{P}$ 和 $\boldsymbol{R}$ 是正定的. 则*

(i) $\boldsymbol{I}(\boldsymbol{P}\,|\,\boldsymbol{R})>0$, $\boldsymbol{P}\neq\boldsymbol{R}$, $\boldsymbol{I}(\boldsymbol{P}\,|\,\boldsymbol{P})=0$.

(ii) *如果 $\{\boldsymbol{P}_n\}$ 和 $\{\boldsymbol{R}_n\}$ 是两个正定矩阵序列, 使得 $\boldsymbol{I}(\boldsymbol{P}_n\,|\,\boldsymbol{R}_n)\to 0$, 则 $\boldsymbol{P}_n\boldsymbol{R}_n^{-1}\to\boldsymbol{I}$.*

证明 (i) 设 $|\boldsymbol{P}-s\boldsymbol{R}|=0$ 的根是 $s_1\leqslant\cdots\leqslant s_p$. 则

$$\ln|\boldsymbol{P}\boldsymbol{R}^{-1}|+\mathrm{tr}\,(\boldsymbol{I}-\boldsymbol{P}\boldsymbol{R}^{-1})=\sum_{i=1}^{p}(\ln s_i+1-s_i)\geqslant 0,\tag{9}$$

并且 (9) 为 0 当且仅当 $s_1=\cdots=s_p=1$.

(ii) 设 $|\boldsymbol{P}_n-s\boldsymbol{R}_n|=0$ 的根是 $s_1(n)\leqslant\cdots\leqslant s_p(n)$. 则 $\boldsymbol{I}(\boldsymbol{P}_n\,|\,\boldsymbol{R}_n)\to 0$ 意味着 $[s_1(n),\cdots,s_p(n)]\to(1,\cdots,1)$, 它意味着 $\boldsymbol{P}_n\boldsymbol{R}_n^{-1}\to\boldsymbol{I}$. ■

引理 15.5.2 设

$$\boldsymbol{P}=\begin{bmatrix}\boldsymbol{P}_{11} & \boldsymbol{P}_{12}\\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22}\end{bmatrix},\qquad \boldsymbol{R}=\begin{bmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12}\\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22}\end{bmatrix}. \tag{10}$$

(i) 矩阵

$$\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{P}_{11} & \boldsymbol{P}_{11}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\\ \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{P}_{11} & \boldsymbol{R}_{22}-\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}+\boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{P}_{11}\boldsymbol{R}_{11}^{-1}\boldsymbol{R}_{12}\end{bmatrix} \tag{11}$$

满足 $\boldsymbol{Q}_{11}=\boldsymbol{P}_{11}$, $\boldsymbol{Q}^{12}=\boldsymbol{R}^{12}$, $\boldsymbol{Q}^{22}=\boldsymbol{R}^{22}$, 其中

$$\boldsymbol{Q}^{-1}=\begin{bmatrix}\boldsymbol{P}_{11}^{-1}+\boldsymbol{R}^{12}(\boldsymbol{R}^{22})^{-1}\boldsymbol{R}^{21} & \boldsymbol{R}^{12}\\ \boldsymbol{R}^{21} & \boldsymbol{R}^{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{P}_{11}^{-1}-\boldsymbol{R}_{11}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}+\boldsymbol{R}^{-1}. \tag{12}$$

(ii) $\boldsymbol{I}(\boldsymbol{P}\,|\,\boldsymbol{R})=\boldsymbol{I}(\boldsymbol{P}\,|\,\boldsymbol{Q})+\boldsymbol{I}(\boldsymbol{Q}\,|\,\boldsymbol{R})$.

证明 (i) 设

$$\boldsymbol{Q}^{-1}=\begin{bmatrix}\boldsymbol{S} & \boldsymbol{R}^{21}\\ \boldsymbol{R}^{21} & \boldsymbol{R}^{22}\end{bmatrix}. \tag{13}$$

则 $\boldsymbol{I}=\boldsymbol{Q}^{-1}\boldsymbol{Q}$ 可解出 $\boldsymbol{S}=\boldsymbol{P}_{11}^{-1}+\boldsymbol{R}^{12}(\boldsymbol{R}^{22})^{-1}\boldsymbol{R}^{21}$, 由定理 A.3.3 得知 $\boldsymbol{Q}=(\boldsymbol{Q}^{-1})^{-1}$. 那么由以下三式推得 (ii):

$$\boldsymbol{P}\boldsymbol{Q}^{-1}=\boldsymbol{P}\boldsymbol{R}^{-1}+\begin{bmatrix}\boldsymbol{I}-\boldsymbol{P}_{11}\boldsymbol{R}_{11}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix},\qquad \boldsymbol{Q}\boldsymbol{R}^{-1}=\begin{bmatrix}\boldsymbol{P}_{11}\boldsymbol{R}_{11}^{-1} & \boldsymbol{0}\\ \boldsymbol{R}_{21}\boldsymbol{R}_{11}^{-1}\boldsymbol{P}_{11} & \boldsymbol{I}\end{bmatrix}, \tag{14}$$

$|\boldsymbol{P}\boldsymbol{Q}^{-1}|\cdot|\boldsymbol{Q}\boldsymbol{R}^{-1}|=|\boldsymbol{P}\boldsymbol{R}^{-1}|$. 根据 (13) 和 (14) 有

$$\operatorname{tr}\boldsymbol{P}\boldsymbol{Q}^{-1}+\operatorname{tr}\boldsymbol{Q}\boldsymbol{R}^{-1}=\operatorname{tr}\boldsymbol{P}\boldsymbol{R}^{-1}+\operatorname{tr}\boldsymbol{I}_p. \qquad \blacksquare \tag{15}$$

引理 15.5.2 提供了对求解矩阵 $\boldsymbol{Q}$ 问题的解, 当给定正定矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{R}$, 使得

$$q_{ij}=p_{ij},\quad (i,j)\in\{1,\cdots,t\}, \tag{16}$$

$$q^{ij}=r^{ij},\quad (i,j)\notin\{1,\cdots,t\}. \tag{17}$$

我们现在发展求满足 (6) 和 (7) 的 $\boldsymbol{K}$ 的一个迭代方法, 于是证明定理 15.5.2. 假设 $E=c_1\cup\cdots\cup c_m$, 其中 $c_1\cdots,c_m$ 是一个可分解图 $G=(V,E)$ 的 m 个团. 设 $\boldsymbol{K}_0^{-1}=\boldsymbol{M}^{-1}$, 递归地定义 $\boldsymbol{K}_n=(k_{ij}(n))$ 使得

$$k_{ij}(n)=l_{ij},\quad i,j\in c_{n\bmod m}, \tag{18}$$

$$k^{ij}(n)=k^{ij}(n-1),\quad i,j\notin c_{n\bmod m}. \tag{19}$$

由引理 15.5.2, $\boldsymbol{K}_n$ 是唯一确定的. (通过团算法循环.) 由构造

$$\boldsymbol{I}(\boldsymbol{L}\,|\,\boldsymbol{K}_{n-1})=\boldsymbol{I}(\boldsymbol{L}\,|\,\boldsymbol{K}_n)+\boldsymbol{I}(\boldsymbol{K}_n\,|\,\boldsymbol{K}_{n-1}). \tag{20}$$

将 (20) 从 1 到 q 求和给出

$$\boldsymbol{I}(\boldsymbol{L}\,|\,\boldsymbol{K}_0)=\boldsymbol{I}(\boldsymbol{L}\,|\,\boldsymbol{K}_q)+\sum_{j=1}^{q}\boldsymbol{I}(\boldsymbol{K}_j\,|\,\boldsymbol{K}_{j-1}). \tag{21}$$

因为 $I(L\,|\,K_q) \geqslant 0$, $\sum_{j=1}^{q} I(K_j\,|\,K_{j-1})$ 是有界的并且当 $n\to\infty$ 时有 $I(L|K_j\,|\,K_{j-1})\to 0$. 集合 $\{K^{-1}\,|\,I(L\,|\,K)\leqslant I(K_0\}$ 是严格凸的.

考虑带指标 r $(n=rm)$ 的向量序列 $K_{rm+1},\cdots,K_{rm+m}$. 它有一个收敛的子序列 $\{r(i)\}$, 即例如 $(K_{mr(i)+1},\cdots,K_{mr(i)+m})$ 收敛于 $(K_1^*,\cdots,K_m^*)$. 因为 $I(K_j\,|\,K_{j-1})\to 0$, 所以 $K_jK_{j-1}^{-1}\to I$. 则矩阵 $K_{mr(i)+j}K_{mr(i)+j-1}^{-1}\to I$, $j=2,\cdots,m$, 其意味着 $K_1^*=\cdots=K_m^*=K$. 注意 $\{i,j\,|\,i,j\notin E\}$ 满足 $i,j\notin c_i, i=1,\cdots,m$. 因此 K_n 满足 (7), $n=0,1,\cdots$, 而且 K 也一样. 进而, $k_{ij}(mr(i)+t)$ 满足 (18), $i,j\in c_t$ 并且 K 也一样. 对于 $c_1=(i,j),i,j=1,\cdots,t$ 和 $c_2=(i,j),i,j=u,u+1,\cdots,p,u<t$, 图 15.8 表示了该集合.

该程序考虑到一个多元正态分布的构造, 这个多元正态分布在团 $c_1,\cdots,c_m$ 上具有任意的边缘分布, 只要指定的边缘分布是相容的.

定理 15.5.2 提供了关于极大似然估计的存在性和唯一性的证明.

方程 (12) 是一个更新方程. 当 $Q^{-1}=K_n^{-1}$ 和 $R^{-1}=K_{n-1}^{-1}$ 时, K_{n-1}^{-1} 的不在 $c_{n \bmod m}$ 中的元素仍然保持不变.

Dempster (1972) 也对求满足 $k_{ij}=l_{ij}[(i,j)\in D]$ 和 $k^{ij}=l^{ij}[(i,j)\notin D]$ 的 K 提出某个迭代方法. $n(x\,|\,0,P)$ 的熵是

$$\mathrm{E}\,(\ln n(x\,|\,\mathbf{0},P)) = -\frac{1}{2}(p\ln 2\pi - p - \ln|P|). \tag{22}$$

注意 $|\Sigma|=\prod_{i=1}^{p}\sigma_{ii}|R|$, 其中 $R=(\rho_{ij})$. 给定 $\hat\sigma_{ii}=s_{ii}$, 极大化满足要求的拟合正态分布的熵 $\hat\rho_{ij}$ 也极小化 $|R|$ [Dempster (1972)].

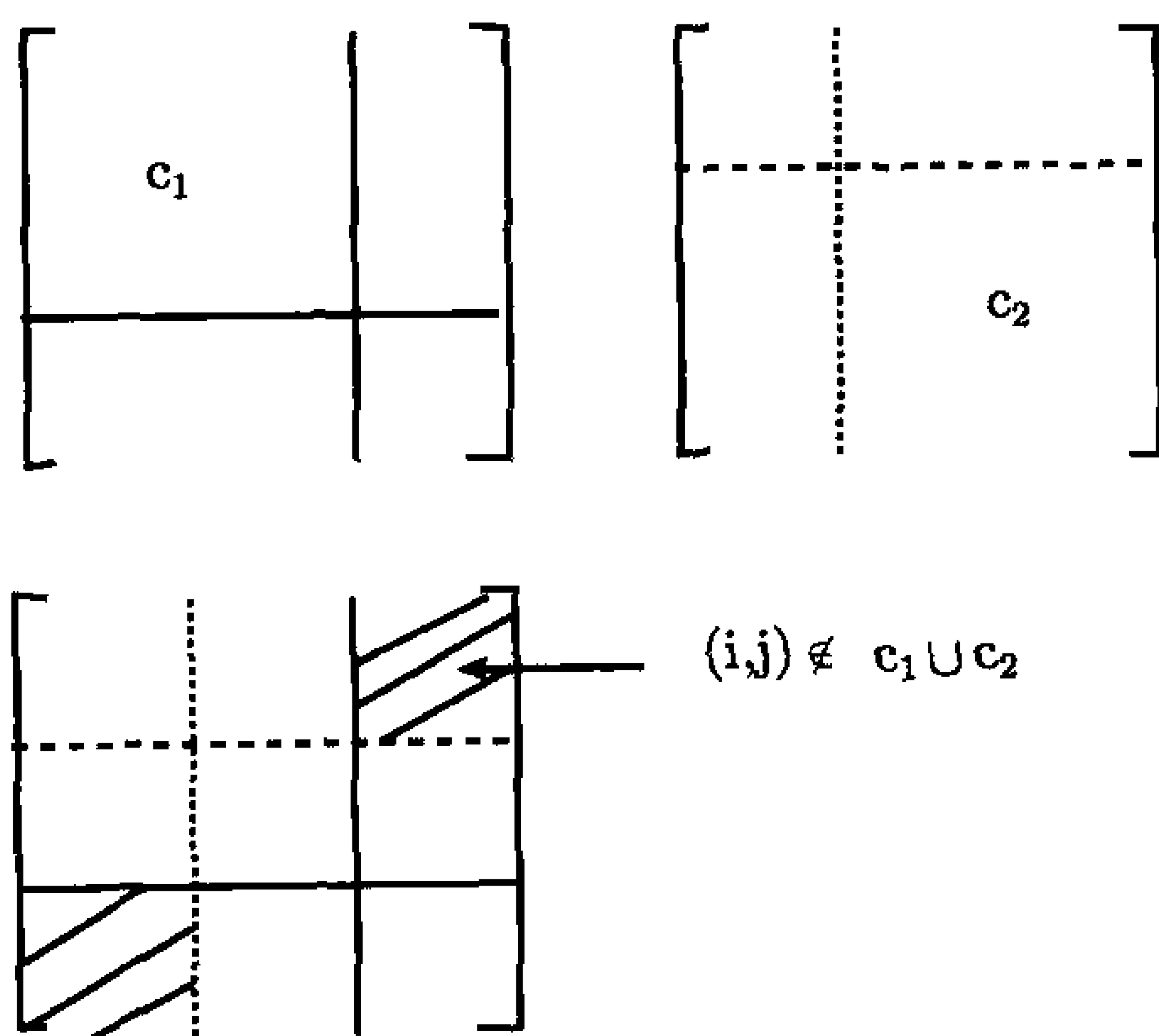


图 15.8 对 $c_1\cup c_2=E$, c_1 和 c_2 的图表示

15.5.3 协方差选择模型的分解

一个无向图是可分解的, 如果该图由三个不相交的集 A,B,C 构成的, 其中

$V = A \cup B \cup C$, A 和 B 是非空的, C 分离 A 和 B, 而且 C 是完全的. 则如果 $\boldsymbol{X}_V$ 关于 G 是整体马氏的, 那么我们有 $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B \mid \boldsymbol{X}_C$,

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_{AA} & \boldsymbol{0} & \boldsymbol{\Lambda}_{AC} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_{BB} & \boldsymbol{\Lambda}_{BC} \\ \boldsymbol{\Lambda}_{CA} & \boldsymbol{\Lambda}_{CB} & \boldsymbol{\Lambda}_{CC} \end{bmatrix}, \tag{23}$$

$$\boldsymbol{\Sigma}_{(AB)\cdot C} = \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Sigma}_{AC} \\ \boldsymbol{\Sigma}_{BC} \end{bmatrix} \boldsymbol{\Sigma}_{CC}^{-1} \begin{bmatrix} \boldsymbol{\Sigma}_{CA} & \boldsymbol{\Sigma}_{CB} \end{bmatrix}, \tag{24}$$

并且

$$\boldsymbol{\Sigma}_{AB} - \boldsymbol{\Sigma}_{AC}\boldsymbol{\Sigma}_{CC}^{-1}\boldsymbol{\Sigma}_{CB} = \boldsymbol{0}. \tag{25}$$

$\boldsymbol{\Sigma}$ 的极大似然估计可以由 $\boldsymbol{\Sigma}_{AA\cdot C}$, $\boldsymbol{\Sigma}_{AB\cdot C}$, $\boldsymbol{\Sigma}_{BB\cdot C}$ 和 $\boldsymbol{\Sigma}_{CC}$ 的极大似然估计来构造.

如果对 $\boldsymbol{\Sigma}$ 没有限制, 则 $\boldsymbol{\Sigma}$ 的极大似然估计是

$$\hat{\boldsymbol{\Sigma}}_{\Omega} = \begin{bmatrix} \boldsymbol{S}_{(AB)\cdot C} + \boldsymbol{S}_{(AB)C}\boldsymbol{S}_{CC}^{-1}\boldsymbol{S}_{C(AB)} & \boldsymbol{S}_{(AB)C} \\ \boldsymbol{S}_{C(AB)} & \boldsymbol{S}_{CC} \end{bmatrix} = \boldsymbol{S}_{ABC}, \tag{26}$$

其中

$$\boldsymbol{S}_{(AB)\cdot C} = \begin{bmatrix} \boldsymbol{S}_{AA\cdot C} & \boldsymbol{S}_{AB\cdot C} \\ \boldsymbol{S}_{BA\cdot C} & \boldsymbol{S}_{BB\cdot C} \end{bmatrix}, \qquad \boldsymbol{S}_{(AB)C} = (\boldsymbol{S}_{AC}, \boldsymbol{S}_{BC}). \tag{27}$$

如果加上限制 (25), 则其极大似然估计就是 (26), 其中 $\boldsymbol{S}_{(AB)\cdot C}$ 用 $\boldsymbol{0}$ 代替而得到

$$\hat{\boldsymbol{\Sigma}}_{\omega} = \begin{bmatrix} \begin{pmatrix} \boldsymbol{S}_{AA\cdot C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{BB\cdot C} \end{pmatrix} + \begin{pmatrix} \boldsymbol{S}_{AC} \\ \boldsymbol{S}_{BC} \end{pmatrix} \boldsymbol{S}_{CC}^{-1}(\boldsymbol{S}_{CA}, \boldsymbol{S}_{CB}) & \begin{pmatrix} \boldsymbol{S}_{AC} \\ \boldsymbol{S}_{BC} \end{pmatrix} \\ (\boldsymbol{S}_{CA}, \boldsymbol{S}_{CB}) & \boldsymbol{S}_{CC} \end{bmatrix}. \tag{28}$$

矩阵 $\boldsymbol{S}_{(AB)\cdot C}$ 有 Wishart 分布 $W[\boldsymbol{\Sigma}_{(AB)\cdot C}, n-(p_A+p_B)]$, 其中 p_A 和 p_B 分别是 $\boldsymbol{X}_A$ 和 $\boldsymbol{X}_B$ 的分量的个数 (8.3 节). 矩阵 $\boldsymbol{B}_{(AB)\cdot C} = \boldsymbol{S}_{(AB)C}\boldsymbol{S}_{CC}^{-1}$ 在条件 $(X_{C1}, \cdots, X_{Cn}) = \boldsymbol{X}_C$ 下有正态分布, 其协方差阵由下式给出,

$$\mathrm{E}\left\{ \mathrm{vec}\begin{bmatrix} \boldsymbol{B}_{A\cdot C} \\ \boldsymbol{B}_{B\cdot C} \end{bmatrix} \left(\mathrm{vec}\begin{bmatrix} \boldsymbol{B}_{A\cdot C} \\ \boldsymbol{B}_{B\cdot C} \end{bmatrix} \right)' \middle| \boldsymbol{S}_{CC} \right\} = \boldsymbol{S}_{CC}^{-1} \otimes \begin{bmatrix} \Sigma_{AA} & \boldsymbol{0} \\ \boldsymbol{0} & \Sigma_{BB} \end{bmatrix}, \tag{29}$$

$\boldsymbol{S}_{CC}$ 有 Wishart 分布 $W(\boldsymbol{\Sigma}_{CC}, n)$. 矩阵 $\boldsymbol{S}_{(AB)\cdot C}$ 和 $\boldsymbol{B}_{(AB)\cdot C}$ 是独立的 (第 8 章).

考虑检验原假设 (25). 这即是检验原假设 $\boldsymbol{\Sigma}_{(AB)\cdot C} = \boldsymbol{0}$ 相对于备择假设 $\boldsymbol{\Sigma}_{(AB)\cdot C} \neq \boldsymbol{0}$. (26) 的行列式是 $|\hat{\boldsymbol{\Sigma}}_{\Omega}| = |\boldsymbol{S}_{AB\cdot C}| \cdot |\boldsymbol{S}_{CC}|$, (26) 的行列式是 $|\hat{\boldsymbol{\Sigma}}_{\omega}| = |\boldsymbol{S}_{AA}| \cdot |\boldsymbol{S}_{BB}| \cdot |\boldsymbol{S}_{CC}|$. 似然比准则是

$$\left(\frac{|\hat{\boldsymbol{\Sigma}}_{\omega}|}{|\hat{\boldsymbol{\Sigma}}_{\Omega}|} \right)^{n/2} = \left(\frac{|\boldsymbol{S}_{(AB)\cdot C}|}{|\boldsymbol{S}_{(AA)\cdot C}| \cdot |\boldsymbol{S}_{(BB)\cdot C}|} \right)^{n/2}. \tag{30}$$

因为样本协方差阵 $\boldsymbol{S}_{(AB)\cdot C}$ 有 Wishart 分布 $W[\boldsymbol{\Sigma}_{(AB)\cdot C}, n-(p_A+p_B)]$, 其中 p_A 和 p_B 分别是 $\boldsymbol{X}_A$ 和 $\boldsymbol{X}_B$ 的分量的个数 (8.2 节), 所以实际上该准则是在 8.4 节和 8.5 节研究的 $u_{p_A, p_B, n-(p_A+p_B)}$.

作为另一个例子, 考虑图 15.9 的图. 注意结 4 分离 (1,2,3) 和 (5,6); 结 1,4 分离 2 和 3; 而结 4 分离 5 和 6. 这些分离意味着三个条件独立性: $(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3) \perp\!\!\!\perp (\boldsymbol{X}_5,\boldsymbol{X}_6) \mid \boldsymbol{X}_4$, $\boldsymbol{X}_2 \perp\!\!\!\perp \boldsymbol{X}_3 \mid (\boldsymbol{X}_1,\boldsymbol{X}_4)$, $\boldsymbol{X}_5 \perp\!\!\!\perp \boldsymbol{X}_6 \mid \boldsymbol{X}_4$. 利用协方差, 这些独立性就是

$$\boldsymbol{\Sigma}_{(123)(56)\cdot 4} = \boldsymbol{\Sigma}_{(123)(56)} - \boldsymbol{\Sigma}_{(123)4}\boldsymbol{\Sigma}_{44}^{-1}\boldsymbol{\Sigma}_{4(56)} = \mathbf{0}, \tag{31}$$

$$\boldsymbol{\Sigma}_{23\cdot(14)} = \boldsymbol{\Sigma}_{23} - \boldsymbol{\Sigma}_{2(14)}\boldsymbol{\Sigma}_{(14)(14)}^{-1}\boldsymbol{\Sigma}_{(14)3} = \mathbf{0}, \tag{32}$$

$$\boldsymbol{\Sigma}_{56\cdot 4} = \boldsymbol{\Sigma}_{56} - \boldsymbol{\Sigma}_{54}\boldsymbol{\Sigma}_{44}^{-1}\boldsymbol{\Sigma}_{46} = \mathbf{0}. \tag{33}$$

从 (31) 看, 限制 (32) 可以写成为

$$\boldsymbol{\Sigma}_{23\cdot 14} = \boldsymbol{\Sigma}_{23\cdot 4} - \boldsymbol{\Sigma}_{21\cdot 4}\boldsymbol{\Sigma}_{11\cdot 4}^{-1}\boldsymbol{\Sigma}_{12\cdot 4} = \mathbf{0}. \tag{34}$$

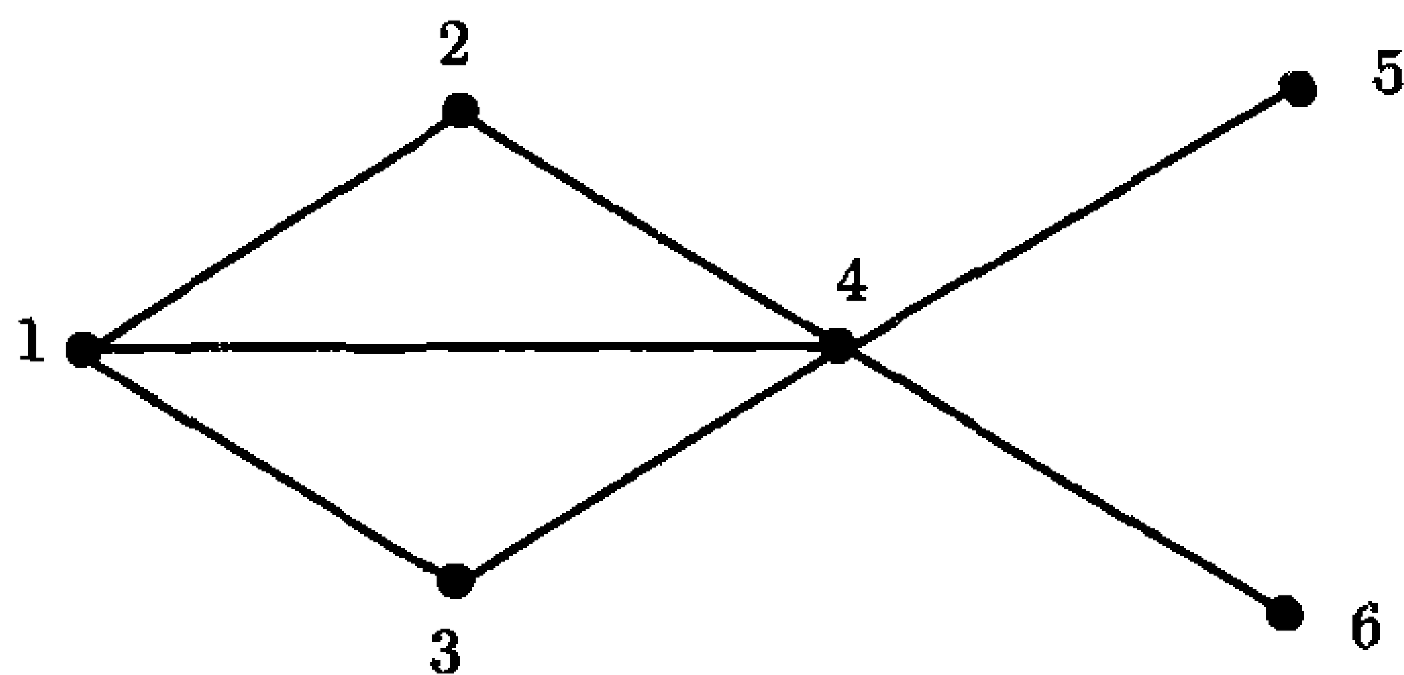


图 15.9

为方便起见, 我们将子这向量重新排序为 $\boldsymbol{X}_2,\boldsymbol{X}_3,\boldsymbol{X}_1,\boldsymbol{X}_5,\boldsymbol{X}_6,\boldsymbol{X}_4$, 便写得

$$\begin{aligned}
\boldsymbol{S} &= \begin{bmatrix} \boldsymbol{S}_{22} & \cdots & \boldsymbol{S}_{26} & \boldsymbol{S}_{24} \\ \vdots & & \vdots & \vdots \\ \boldsymbol{S}_{62} & \cdots & \boldsymbol{S}_{66} & \boldsymbol{S}_{64} \\ \boldsymbol{S}_{42} & \cdots & \boldsymbol{S}_{46} & \boldsymbol{S}_{44} \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} \boldsymbol{S}_{22\cdot 4} & \cdots & \boldsymbol{S}_{26\cdot 4} \\ \vdots & & \vdots \\ \boldsymbol{S}_{62\cdot 4} & \cdots & \boldsymbol{S}_{66\cdot 4} \end{bmatrix} + \begin{bmatrix} \boldsymbol{S}_{24} \\ \vdots \\ \boldsymbol{S}_{64} \end{bmatrix} \boldsymbol{S}_{44}^{-1} [\boldsymbol{S}_{42} \ \cdots \ \boldsymbol{S}_{46}] & \begin{bmatrix} \boldsymbol{S}_{24} \\ \vdots \\ \boldsymbol{S}_{64} \end{bmatrix} \\ [\boldsymbol{S}_{42} \ \cdots \ \boldsymbol{S}_{46}] & \boldsymbol{S}_{44} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{S}_{(2\cdots 6)(2\cdots 6)\cdot 4} + \boldsymbol{S}_{(2\cdots 6)4}\boldsymbol{S}_{44}^{-1}\boldsymbol{S}_{4(2\cdots 6)} & \boldsymbol{S}_{(2\cdots 6)4} \\ \boldsymbol{S}_{4(2\cdots 6)} & \boldsymbol{S}_{44} \end{bmatrix}.
\end{aligned} \tag{35}$$

$\boldsymbol{S}$ 的行列式是

$$|\boldsymbol{S}| = |\boldsymbol{S}_{(2\cdots 6)(2\cdots 6)\cdot 4}| \cdot |\boldsymbol{S}_{44}|. \tag{36}$$

如果加上条件 $(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3) \perp\!\!\!\perp (\boldsymbol{X}_5,\boldsymbol{X}_6) \mid \boldsymbol{X}_4$, 则极大似然估计是 (35), 其中用0代替 $\boldsymbol{S}_{(125)(56)\cdot 4}$, 便得

$$\begin{bmatrix} \begin{pmatrix} \boldsymbol{S}_{(231)(231)\cdot 4} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{(56)(56)\cdot 4} \end{pmatrix} + \boldsymbol{S}_{(2\cdots 6)4}\boldsymbol{S}_{44}^{-1}\boldsymbol{S}_{4(2\cdots 6)} & \boldsymbol{S}_{(2\cdots 6)4} \\ \boldsymbol{S}_{4(2\cdots 6)} & \boldsymbol{S}_{44} \end{bmatrix}. \tag{37}$$

(37) 的行列式是

$$|\boldsymbol{S}_{(231)(231)\cdot 4}| = |\boldsymbol{S}_{(56)(56)\cdot 4}| \cdot |\boldsymbol{S}_{44}|. \tag{38}$$

对 $\boldsymbol{\Sigma}_{(125)(56)\cdot 4} = \boldsymbol{0}$ 的似然比准则是

$$\left(\frac{|\boldsymbol{S}_{(2\cdots 6)(2\cdots 6)\cdot 4}|}{|\boldsymbol{S}_{(231)(231)\cdot 4}| \cdot |\boldsymbol{S}_{(56)(56)\cdot 4}|} \right)^{n/2} = U^{n/2}_{(231)(56)\cdot 4}. \tag{39}$$

这里的 $U_{(231)(56)\cdot 4}$ 有分布 $u_{p_1+p_2+p_3, p_5+p_6, n-p}$ (参见 8.4 节), 因为 $\boldsymbol{S}_{(2\cdots 6)(2\cdots 6)\cdot 4}$ 的分布是 $W(\boldsymbol{\Sigma}_{(2\cdots 6)(2\cdots 6)\cdot 4}, n-p_4)$, 与 $\boldsymbol{S}_{44}$ 独立.

$\boldsymbol{S}_{(2\cdots 6)(2\cdots 6)\cdot 4}$ 的前三行和前三列构成矩阵

$$\begin{aligned}
\boldsymbol{S}_{(231)(231)\cdot 4} &= \begin{bmatrix} \boldsymbol{S}_{22\cdot 4} & \boldsymbol{S}_{23\cdot 4} & \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{32\cdot 4} & \boldsymbol{S}_{33\cdot 4} & \boldsymbol{S}_{31\cdot 4} \\ \boldsymbol{S}_{12\cdot 4} & \boldsymbol{S}_{13\cdot 4} & \boldsymbol{S}_{11\cdot 4} \end{bmatrix} \\
&= \begin{bmatrix} \begin{pmatrix} \boldsymbol{S}_{22\cdot 14} & \boldsymbol{S}_{23\cdot 14} \\ \boldsymbol{S}_{32\cdot 14} & \boldsymbol{S}_{33\cdot 14} \end{pmatrix} + \begin{pmatrix} \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{31\cdot 4} \end{pmatrix} \boldsymbol{S}^{-1}_{11\cdot 4} (\boldsymbol{S}_{12\cdot 4}, \boldsymbol{S}_{13\cdot 4}) & \begin{pmatrix} \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{31\cdot 4} \end{pmatrix} \\ (\boldsymbol{S}_{12\cdot 4} \ \cdots \ \boldsymbol{S}_{13\cdot 4}) & \boldsymbol{S}_{11\cdot 4} \end{bmatrix} \\
&= \begin{bmatrix} \begin{pmatrix} \boldsymbol{S}_{22\cdot 14} & \boldsymbol{S}_{23\cdot 14} \\ \boldsymbol{S}_{32\cdot 14} & \boldsymbol{S}_{33\cdot 14} \end{pmatrix} + \begin{pmatrix} \boldsymbol{S}_{2(14)} \\ \boldsymbol{S}_{3(14)} \end{pmatrix} \boldsymbol{S}^{-1}_{(14)(14)} (\boldsymbol{S}_{(14)2}, \boldsymbol{S}_{(14)3}) & \begin{pmatrix} \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{31\cdot 4} \end{pmatrix} \\ (\boldsymbol{S}_{12\cdot 4}, \boldsymbol{S}_{13\cdot 4}) & \boldsymbol{S}_{11\cdot 4} \end{bmatrix}
\end{aligned} \tag{40}$$

(40) 的行列式为

$$|\boldsymbol{S}_{(231)(231)\cdot 4}| = |\boldsymbol{S}_{(23)(23)\cdot 14}| \cdot |\boldsymbol{S}_{11\cdot 4}|. \tag{41}$$

在加约束 $\boldsymbol{X}_2 \amalg \boldsymbol{X}_3 \mid \boldsymbol{X}_1, \boldsymbol{X}_4$ 下, $\boldsymbol{\Sigma}_{(231)(231)\cdot 4}$ 的估计是 (40), 但 $\boldsymbol{S}_{23\cdot 14}$ 用 $\boldsymbol{0}$ 代替, 便得到

$$\begin{bmatrix} \begin{pmatrix} \boldsymbol{S}_{22\cdot 14} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{33\cdot 14} \end{pmatrix} + \begin{pmatrix} \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{31\cdot 4} \end{pmatrix} \boldsymbol{S}^{-1}_{11\cdot 4} (\boldsymbol{S}_{12\cdot 4}, \boldsymbol{S}_{13\cdot 4}) & \begin{pmatrix} \boldsymbol{S}_{21\cdot 4} \\ \boldsymbol{S}_{31\cdot 4} \end{pmatrix} \\ (\boldsymbol{S}_{12\cdot 4}, \boldsymbol{S}_{13\cdot 4}) & \boldsymbol{S}_{11\cdot 4} \end{bmatrix}, \tag{42}$$

它的行列式是

$$|\boldsymbol{S}_{22\cdot 14}| \cdot |\boldsymbol{S}_{33\cdot 14}| \cdot |\boldsymbol{S}_{11\cdot 4}|. \tag{43}$$

对检验 $\boldsymbol{\Sigma}_{23\cdot 14} = \boldsymbol{0}$ 的似然比准则是

$$\left(\frac{|\boldsymbol{S}_{(23)(23)\cdot 14}|}{|\boldsymbol{S}_{22\cdot 14}| \cdot |\boldsymbol{S}_{33\cdot 14}|} \right)^{n/2} = U^{n/2}_{23\cdot 14}. \tag{44}$$

统计量 $U_{23\cdot 14}$ 有 $u_{p_2, p_3, n-(p_1+p_2+p_3+p_4)}$ 的分布 (8.4 节), 因为 $\boldsymbol{S}_{(23)(23)\cdot 14}$ 有与 $\boldsymbol{S}_{(14)(14)}$ 独立的 $W(\boldsymbol{\Sigma}_{(23)(23)\cdot 14}, n-(p_1+p_4))$ 分布.

$\boldsymbol{\Sigma}_{(56)(56)\cdot 4}$ 带约束 $\boldsymbol{\Sigma}_{(56)\cdot 4} = \boldsymbol{0}$ 的估计是 $\boldsymbol{S}_{(56)(56)\cdot 4}$, 其中 $\boldsymbol{S}_{56\cdot 4}$ 用 $\boldsymbol{0}$ 代替, 得到

$$\begin{bmatrix} \boldsymbol{S}_{55\cdot 4} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{66\cdot 4} \end{bmatrix}. \tag{45}$$

对检验 $\boldsymbol{\Sigma}_{56\cdot 4}=\boldsymbol{0}$ 的似然比准则是

$$\left(\frac{|\boldsymbol{S}_{(56)(56)\cdot 4}|}{|\boldsymbol{S}_{55\cdot 4}|\cdot|\boldsymbol{S}_{66\cdot 4}|}\right)^{n/2}=U_{56\cdot 4}^{n/2}. \tag{46}$$

统计量 $U_{56\cdot 4}$ 有 $u_{p_5,p_6,n-(p_4+p_5+p_6)}$ 的分布, 因为 $\boldsymbol{S}_{(56)(56)\cdot 4}$ 有与 $\boldsymbol{S}_{44}$ 独立的 $W(\boldsymbol{\Sigma}_{(56)(56)\cdot 4}, n-p_4)$ 分布.

在三个原假设下 $\boldsymbol{\Sigma}$ 的估计是 (37), 其中的 $\boldsymbol{S}_{(231)(231)\cdot 4}$ 用 (42) 代替而 $\boldsymbol{S}_{(56)(56)\cdot 4}$ 用 (45) 代替. 这个矩阵的行列式是

$$|\hat{\boldsymbol{\Sigma}}_\omega|=|\boldsymbol{S}_{22\cdot 14}|\cdot|\boldsymbol{S}_{33\cdot 14}|\cdot|\boldsymbol{S}_{11\cdot 4}|\cdot|\boldsymbol{S}_{55\cdot 4}|\cdot|\boldsymbol{S}_{66\cdot 4}|\cdot|\boldsymbol{S}_{44}|. \tag{47}$$

对检验这三个原假设的似然比准则是

$$\left(\frac{|\hat{\boldsymbol{\Sigma}}_\Omega|}{|\hat{\boldsymbol{\Sigma}}_\omega|}\right)^{n/2}=\left(U_{(231)(56)\cdot 4}U_{23\cdot 14}U_{56\cdot 4}\right)^{n/2}. \tag{48}$$

当这三个原假设为真, 因子 $U_{(231)(56)\cdot 4}$,$U_{23\cdot 14}$ 和 $U_{56\cdot 4}$ 是独立的. 它们的分布在 8.4 和 8.5 节有讨论. 特别地, 这些因子的矩被给出并且分布的渐近展开也被描述.

15.5.4 有向图

我们假设顶点是完好编号的, $1,\cdots,n$, 在 $\boldsymbol{X}=(\boldsymbol{X}_1',\cdots,\boldsymbol{X}_p')'$ 上有 N 个观测 $\boldsymbol{x}_{(1)},\cdots,\boldsymbol{x}_{(N)}$. 模型是 15.3 节的从 (22) 到 (25) 或 (26). 设 $\bar{\boldsymbol{x}}=N^{-1}\sum_{\alpha=1}^N\boldsymbol{x}_{(\alpha)}$ 和 $\boldsymbol{S}=(N-1)^{-1}\sum_{\alpha=1}^N(\boldsymbol{x}_{(\alpha)}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{(\alpha)}-\bar{\boldsymbol{x}})'$. 模型 (26) 是由 $\boldsymbol{x}_1=\boldsymbol{\alpha}_1+\boldsymbol{\varepsilon}_1$ 和 $n-1$ 个从 (23) 到 (25) 的回归组成. 在 $\boldsymbol{x}_1=\boldsymbol{\alpha}_1+\boldsymbol{\varepsilon}_1$ 中的向量 $\boldsymbol{\alpha}_1$ 由 $\bar{x}_1$ 估计. 如果 $\mathrm{pa}(v_2)$ 是空的, 则 $\boldsymbol{\alpha}_2$ 由 $\bar{x}_2$ 估计; 如果 $\mathrm{pa}(v_2)$ 不空并且 $\boldsymbol{X}_{\mathrm{pa}(v_2)}=\boldsymbol{X}_1$, 则 $\boldsymbol{B}_2$ 和 $\boldsymbol{\alpha}_2$ 由下面两式估计:

$$\hat{\boldsymbol{B}}_2=\sum_{\alpha=1}^N(\boldsymbol{x}_{2(\alpha)}-\bar{\boldsymbol{x}}_2)(\boldsymbol{x}_{1(\alpha)}-\bar{\boldsymbol{x}}_1)'\left[\sum_{\alpha=1}^N(\boldsymbol{x}_{1(\alpha)}-\bar{\boldsymbol{x}}_1)(\boldsymbol{x}_{1(\alpha)}-\bar{\boldsymbol{x}}_1)'\right]^{-1} \tag{49}$$

$$\hat{\boldsymbol{\alpha}}_2=\bar{\boldsymbol{x}}_2+\hat{\boldsymbol{B}}_2\bar{\boldsymbol{x}}_1. \tag{50}$$

一般地,

$$\hat{\boldsymbol{B}}_j=\sum_{\alpha=1}^N[\boldsymbol{x}_{j(\alpha)}-\bar{\boldsymbol{x}}_j][\boldsymbol{x}_{\mathrm{pa}(v_j)(\alpha)}-\bar{\boldsymbol{x}}_{\mathrm{pa}(v_j)}]' \tag{51}$$

$$\cdot\left\{\sum_{\alpha=1}^N[\boldsymbol{x}_{\mathrm{pa}(v_j)(\alpha)}-\bar{\boldsymbol{x}}_{\mathrm{pa}(v_j)}][\boldsymbol{x}_{\mathrm{pa}(v_j)(\alpha)}-\bar{\boldsymbol{x}}_{\mathrm{pa}(v_j)}]'\right\}^{-1},$$

$$\hat{\boldsymbol{\alpha}}_j=\bar{\boldsymbol{x}}_j+\hat{\boldsymbol{B}}_j\bar{\boldsymbol{x}}_{\mathrm{pa}(v_j)}. \tag{52}$$

给定 $\boldsymbol{x}_{\mathrm{pa}(v_j)(\alpha)}$ 的条件下, 这些分布是正态的.

15.5.5 链图

15.4 节的条件 (C1) 对 $u \in V(\tau)$ 和 $v \in V(\sigma)$ 指定 $\boldsymbol{X}_u \amalg \boldsymbol{X}_v \mid \boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$, 其中 $\sigma \in \mathrm{nd}_{\mathscr{D}}(\tau) \setminus \mathrm{pa}_{\mathscr{D}}(\tau)$, 即比 $\mathrm{pa}_{\mathscr{D}}(\tau)$ 早的过去与现在独立. 这个条件对应着在时间序列分析中的马氏性. 于是 $\boldsymbol{X}_u$ 用 $\boldsymbol{X}_\tau$ 关于 $\boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$ 回归的偏差来表示的,

$$\mathrm{E}(\boldsymbol{X}_\tau \mid \boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}) = \boldsymbol{\alpha}_\tau + \boldsymbol{B}_\tau \boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}.$$

向量 $\boldsymbol{\alpha}_\tau$ 和矩阵 $\boldsymbol{B}_\tau$ 像对有向图那样被估计.

马氏性 (C2) 表明了用偏差 $\boldsymbol{X}_\tau - \boldsymbol{\alpha}_\tau + \boldsymbol{B}_\tau \boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$ 的分析. 在 $V(\tau)$ 内相依性的结构的估计如在 15.5.2 节中那样进行.

马氏性 (C3*) 对 $u \in U \subseteq V(\tau)$ 和 $v \in \mathrm{pa}_{\mathscr{D}}(U) \cup \mathrm{nb}_G(U)$ 指定 $\boldsymbol{X}_u \amalg \boldsymbol{X}_v \mid \boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$. 该性质是对 $\boldsymbol{X}_\tau$ 在 $\boldsymbol{X}_{\mathrm{pa}_{\mathscr{D}}(\tau)}$ 上回归的限制.

致谢

本章的写作主要利用了 Michael Perlman 的演讲稿. 同时还感谢 Michael Perlman 和 Ingram Olkin 阅读这章的初稿.

附录A　矩 阵 理 论

A.1　矩阵和矩阵运算的定义

在本附录中我们概括介绍矩阵代数的一些著名的定义和定理. 许多结果并不总是出现在关于矩阵代数的书中, 在这里给出了证明.

一个 $m \times n$ 矩阵$\boldsymbol{A}$是一个实数长方阵列

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \tag{1}$$

它可以被简写成 (a_{ij}), $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, n$. 我们用黑体的大写字母来表示矩阵, 而它的元素用相对应的小写字母加上适当的下标来表示. 两个分别具有相同行数和列数的矩阵$\boldsymbol{A}$和矩阵$\boldsymbol{B}$的和, 定义为

$$\boldsymbol{A} + \boldsymbol{B} = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}). \tag{2}$$

实数 λ 与一个矩阵相乘定义为

$$\lambda \boldsymbol{A} = \boldsymbol{A} \lambda = (\lambda a_{ij}). \tag{3}$$

这些运算具有代数性质

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}, \tag{4}$$

$$(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}), \tag{5}$$

$$\boldsymbol{A} + (-1)\boldsymbol{A} = (0), \tag{6}$$

$$(\lambda + \mu)\boldsymbol{A} = \lambda \boldsymbol{A} + \mu \boldsymbol{A}, \tag{7}$$

$$\lambda(\boldsymbol{A} + \boldsymbol{B}) = \lambda \boldsymbol{A} + \lambda \boldsymbol{B}, \tag{8}$$

$$\lambda(\mu \boldsymbol{A}) = (\lambda \mu)\boldsymbol{A}. \tag{9}$$

所有元素都为 0 的矩阵 (0) 记作$\boldsymbol{0}$. 运算 $\boldsymbol{A} + (-1)\boldsymbol{B}$ 记作 $\boldsymbol{A} - \boldsymbol{B}$.

如果$\boldsymbol{A}$的列数与$\boldsymbol{B}$的行数相同, 也就是说, $\boldsymbol{A} = (a_{ij})$, $i = 1, \cdots, l$, $j = 1, \cdots, m$, $\boldsymbol{B} = (b_{jk})$, $j = 1, \cdots, m$, $k = 1, \cdots, n$, 那么$\boldsymbol{A}$与$\boldsymbol{B}$可以按照下面的法则相乘,

$$\boldsymbol{A}\boldsymbol{B} = (a_{ij})(b_{jk}) = \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right), \quad i = 1, \cdots, l, \ k = 1, \cdots, n. \tag{10}$$

也就是说, $\boldsymbol{A}\boldsymbol{B}$ 是一个具有 l 行和 n 列的矩阵, 它的第 i 行第 k 列的元素是 $\sum_{j=1}^{m} a_{ij} b_{jk}$. 矩阵的乘法具有下面的性质,

$$(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}), \tag{11}$$

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\boldsymbol{C}, \tag{12}$$

$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{A}\boldsymbol{C} + \boldsymbol{B}\boldsymbol{C}. \tag{13}$$

当一侧有意义时 (即矩阵的行数和列数使得矩阵的乘法能够进行), 关系式 (11)~(13) 成立; 当另一侧有意义时结论也同样成立. 因为 (11) 式, 我们可以写作

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) = \boldsymbol{ABC}. \tag{14}$$

乘积$\boldsymbol{BA}$可能是没有意义的, 即使$\boldsymbol{AB}$是有意义的; 甚至当它们都有意义时, 它们也不一定相等.

一个 $l \times m$ 矩阵 $\boldsymbol{A} = (a_{ij})$ 的**转置**定义为矩阵 $\boldsymbol{A}'$, 其第 j 行第 i 列的元素恰是矩阵$\boldsymbol{A}$的第 i 行第 j 列的元素. 转置运算具有下面的性质

$$(\boldsymbol{A}')' = \boldsymbol{A}, \tag{15}$$

$$(\boldsymbol{A}+\boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}', \tag{16}$$

$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}', \tag{17}$$

这里同样要有一个限制条件, 即至少有一侧是有意义的, 这个限制条件贯穿全书.

一个具有 m 个分量的向量$\boldsymbol{x}$可以当作一个 m 行 1 列的矩阵. 因此, 以上的运算对向量同样适用.

现在, 我们将讨论具有同样大小的方阵, 它们可以任意地相加或相乘. 假定行数和列数均为 p. 称方阵$\boldsymbol{A}$为是对称的, 如果 $\boldsymbol{A} = \boldsymbol{A}'$. 一个值得关心的特殊矩阵是**单位矩阵**

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = (\delta_{ij}), \tag{18}$$

其中, δ_{ij} (Kronecker delta) 定义为

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{19}$$

单位矩阵满足

$$\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}. \tag{20}$$

当我们想要强调单位矩阵的阶数为 p 阶时, 应该把它写作$\boldsymbol{I}_p$. 任何一个方阵$\boldsymbol{A}$ 的行列式 $|\boldsymbol{A}|$ 定义为

$$|\boldsymbol{A}| = \sum (-1)^{f(j_1,\cdots,j_p)} \prod_{i=1}^{p} a_{ij_i}, \tag{21}$$

这里的求和是针对整数集合 $(1,\cdots,p)$ 的所有排列 $(j_1,\cdots,j_p)$ 进行的, 而 $f(j_1,\cdots,j_p)$ 是把 $(1,\cdots,p)$ 变成 $(j_1,\cdots,j_p)$ 所需要的对换的数目. 一个对换是由交换两个数字所组成的, 可以证明, 尽管用对换把 $(1,\cdots,p)$ 变成 $(j_1,\cdots,j_p)$ 的方式有很多, 但是所需要对换数总是偶数或者总是奇数, 因此,$(-1)^{f(j_1,\cdots,j_p)}$ 始终会被定义. 那么, 就有

$$|\boldsymbol{AB}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|. \tag{22}$$

也有

$$|\boldsymbol{A}| = |\boldsymbol{A}'|. \tag{23}$$

矩阵$\boldsymbol{A}$的**子矩阵**是通过删除$\boldsymbol{A}$的行和列所得到长方阵列. **子式**是$\boldsymbol{A}$的一个子方阵的行列式. 元素 a_{ij} 的子式是删除矩阵$\boldsymbol{A}$的第 i 行和第 j 列后所得到的子矩阵的行列式. 元素 a_{ij} 的**代数余子式**是 $(-1)^{i+j}$ 乘以 a_{ij} 的子式, 记作 A_{ij}. 由 (21) 式可得

$$|\boldsymbol{A}| = \sum_{j=1}^{p} a_{ij}A_{ij} = \sum_{j=1}^{p} a_{jk}A_{jk}. \tag{24}$$

如果 $|\boldsymbol{A}| \neq 0$, 存在唯一的矩阵 $\boldsymbol{B}$, 使得 $\boldsymbol{AB}=\boldsymbol{I}$. 那么 $\boldsymbol{B}$ 称为 $\boldsymbol{A}$ 的*逆矩阵*, 记作 $\boldsymbol{A}^{-1}$. 设 $\boldsymbol{A}^{-1}$ 的第 h 行第 k 列的元素为 a^{hk}. 那么

$$a^{hk}=\frac{A_{kh}}{|\boldsymbol{A}|}. \tag{25}$$

求逆运算满足

$$(\boldsymbol{AC})^{-1}=\boldsymbol{C}^{-1}\boldsymbol{A}^{-1}, \tag{26}$$

因为

$$(\boldsymbol{AC})(\boldsymbol{C}^{-1}\boldsymbol{A}^{-1})=\boldsymbol{A}(\boldsymbol{CC}^{-1})\boldsymbol{A}^{-1}=\boldsymbol{AIA}^{-1}=\boldsymbol{AA}^{-1}=\boldsymbol{I}. \tag{27}$$

同样 $\boldsymbol{I}^{-1}=\boldsymbol{I}$, 而且 $\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}$. 再进一步, 由于 (27) 式的转置给出 $(\boldsymbol{A}^{-1})'\boldsymbol{A}'=\boldsymbol{I}$, 我们得到 $(\boldsymbol{A}^{-1})'=(\boldsymbol{A})'^{-1}$.

如果一个矩阵的行列式不为零, 那么称它为*非奇异的*. 如果 $|\boldsymbol{A}| \neq 0$, 那么

$$\boldsymbol{Az}=\boldsymbol{0} \tag{28}$$

的唯一解是平凡解 $\boldsymbol{z}=\boldsymbol{0}$ [在 (28) 式两端同时左乘 $\boldsymbol{A}^{-1}$ 即可得到]. 如果 $|\boldsymbol{A}|=0$, 那么至少有一个非平凡解 (即 $\boldsymbol{z}\neq\boldsymbol{0}$). 因此, 矩阵$\boldsymbol{A}$是非奇异矩阵的一个等价定义为, (28) 式只有平凡解.

一组向量 $\boldsymbol{z}_1,\cdots,\boldsymbol{z}_r$ 称为是*线性无关的*, 如果不存在一组不全为零的常数 $c_1,\cdots,c_r$, 使得 $\sum_{i=1}^{r}c_i\boldsymbol{z}_i=\boldsymbol{0}$. 称一个 $q\times p$ 矩阵 $\boldsymbol{D}$ 的*秩*为 r, 如果它的最大的线性无关的列数为 r. 这样, 每一个 $r+1$ 阶子式都必为 0 (将前面讨论的结果应用于相应的 $r+1$ 阶方阵可得), 而且至少有一个 r 阶子式不为 0. 反之, 如果至少有一个 r 阶子式不为 0, 那么至少有一组由 r 个列 (或者行) 组成的向量组是线性无关的. 如果所有的 $r+1$ 阶子式都为 0, 那么将不会存在一组由 $r+1$ 个列 (或者行) 组成的向量组是线性无关的, 因为这样的线性无关将意味着存在一个 $r+1$ 阶的非零子式, 但这与假设相矛盾. 因此, 秩 r 可以等价地由最大的线性无关的行数, 或者由最大的线性无关的列数, 或者由最大的非零子式的阶数来定义.

我们现在考虑如下的二次型

$$\boldsymbol{x}'\boldsymbol{Ax}=\sum_{i,j=1}^{p}a_{ij}x_ix_j, \tag{29}$$

其中,$\boldsymbol{x}'=(x_1,\cdots,x_p)$, 而 $\boldsymbol{A}=(a_{ij})$ 是对称矩阵. 矩阵$\boldsymbol{A}$与二次型称为是*半正定的*, 如果对于所有的 $\boldsymbol{x}$ 都有 $\boldsymbol{x}'\boldsymbol{Ax}\geqslant 0$. 如果 $\boldsymbol{x}'\boldsymbol{Ax}>0$ 对于所有的 $\boldsymbol{x}\neq\boldsymbol{0}$ 都成立, 那么矩阵 $\boldsymbol{A}$ 与二次型称为是*正定的*. 在本书中, 正定就意味着矩阵是对称的.

定理 A.1.1 *如果 $\boldsymbol{C}$ 是一个 p 行 p 列的正定矩阵, $\boldsymbol{B}$ 是 p 行 q 列的矩阵 $(q\leqslant p)$, 秩为 q. 那么 $\boldsymbol{B}'\boldsymbol{CB}$ 是正定的.*

证明 任意给定一个向量 $\boldsymbol{y}\neq\boldsymbol{0}$, 令 $\boldsymbol{x}=\boldsymbol{By}$. 因为 $\boldsymbol{B}$ 的秩为 q, 所以 $\boldsymbol{By}=\boldsymbol{x}\neq\boldsymbol{0}$. 那么

$$\begin{aligned}\boldsymbol{y}'(\boldsymbol{B}'\boldsymbol{CB})\boldsymbol{y}&=(\boldsymbol{By})'\boldsymbol{C}(\boldsymbol{By})\\&=\boldsymbol{x}'\boldsymbol{Cx}>0.\end{aligned} \tag{30}$$

我们看到 $\boldsymbol{B}'\boldsymbol{CB}$ 是对称的, 从而证明完成. 作为逆命题, 我们发现只有当 $\boldsymbol{B}$ 的秩为 q 时 $\boldsymbol{B}'\boldsymbol{CB}$ 是正定的, 否则, 存在 $\boldsymbol{y}\neq\boldsymbol{0}$ 使得 $\boldsymbol{By}=\boldsymbol{0}$. ■

推论 A.1.1 *如果$\boldsymbol{C}$是正定的, $\boldsymbol{B}$ 是非奇异的, 那么$\boldsymbol{B}'\boldsymbol{CB}$是正定的.*

推论 A.1.2 *如果 $\boldsymbol{C}$ 是正定的, 则 $\boldsymbol{C}^{-1}$ 也是正定的.*

证明 C 一定是非奇异的, 因为如果对于某一个 $\boldsymbol{x} \neq \mathbf{0}$ 有 $\boldsymbol{C}\boldsymbol{x} = \mathbf{0}$, 那么对于这个 $\boldsymbol{x}$, 就有 $\boldsymbol{x}'\boldsymbol{C}\boldsymbol{x} = 0$, 但是这与假设 $\boldsymbol{C}$ 是正定的相矛盾. 令 $\boldsymbol{C}^{-1}$ 代替定理 A.1.1 中的 $\boldsymbol{B}$. 那么 $\boldsymbol{B}'\boldsymbol{C}\boldsymbol{B} = (\boldsymbol{C}^{-1})'\boldsymbol{C}\boldsymbol{C}^{-1} = (\boldsymbol{C}^{-1})'$. 对 $\boldsymbol{C}\boldsymbol{C}^{-1} = \boldsymbol{I}$ 求转置, 我们有 $(\boldsymbol{C}^{-1})'\boldsymbol{C}' = (\boldsymbol{C}^{-1})'\boldsymbol{C} = \boldsymbol{I}$. 因此, $\boldsymbol{C}^{-1} = (\boldsymbol{C}^{-1})'$. ■

推论 A.1.3 从一个正定矩阵 $\boldsymbol{C}$ 中删除它的 $p-q$ 个行, 以及相应的 $p-q$ 个列, 所得到的 $q \times q$ 矩阵是正定的.

证明 对应从矩阵 $\boldsymbol{C}$ 中所删除的列, 在 $p \times p$ 单位矩阵中删除相应的列, 得到矩阵 $\boldsymbol{B}$. 利用定理 A.1.1 的结果, 即可得到这个推论. ■

方阵 $\boldsymbol{A}$ 的迹定义为 $\mathrm{tr}\boldsymbol{A} = \sum_{i=1}^{p} a_{ii}$. 下面的性质可以直接得到验证:

$$\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}\boldsymbol{A} + \mathrm{tr}\boldsymbol{B}, \tag{31}$$

$$\mathrm{tr}\boldsymbol{A}\boldsymbol{B} = \mathrm{tr}\boldsymbol{B}\boldsymbol{A}. \tag{32}$$

称方阵 $\boldsymbol{A}$ 为*对角形的*, 如果 $a_{ij} = 0$, $i \neq j$. 这样, 由于在 (24) 式中有 $|\boldsymbol{A}| = a_{11}A_{11}$, $|\boldsymbol{A}| = \prod_{i=1}^{p} a_{ii}$, 同样, A_{11} 有相似的结果.

称方阵 $\boldsymbol{A}$ 为*三角形的*, 如果 $a_{ij} = 0$, 对于 $i > j$ 或者 $i < j$. 如果对于 $i > j$ 有 $a_{ij} = 0$, 这个矩阵称为*上三角矩阵*, 而如果对于 $i < j$ 有 $a_{ij} = 0$, 则称为*下三角矩阵*. 两个上三角矩阵$\boldsymbol{A}$,$\boldsymbol{B}$的乘积是上三角的, 由于 $\boldsymbol{A}\boldsymbol{B}$ 的 (i, j) 元 $(i > j)$ 是 $\sum_{k=1}^{p} a_{ik}b_{kj} = 0$, 因为当 $k < i$ 时有 $a_{ik} = 0$, 当 $k > j$ 时有 $b_{kj} = 0$. 同样, 两个下三角矩阵的乘积是下三角的. 三角形矩阵的行列式是对角线上元素的乘积. 同理, 一个非奇异的三角形矩阵的逆矩阵同样是三角形的.

定理 A.1.2 如果 $\boldsymbol{A}$ 是非奇异矩阵, 那么存在一个非奇异的下三角矩阵$\boldsymbol{F}$ 使得 $\boldsymbol{F}\boldsymbol{A} = \boldsymbol{A}^*$ 是非奇异的上三角矩阵.

证明 令 $\boldsymbol{A} = \boldsymbol{A}_1$. 递归地定义 $\boldsymbol{A}_g = (a_{ij}^{(g)}) = \boldsymbol{F}_{g-1}\boldsymbol{A}_{g-1}$, $g = 2, \cdots, p$, 其中 $\boldsymbol{F}_{g-1} = (f_{ij}^{(g-1)})$ 的元为

$$f_{jj}^{(g-1)} = 1, \qquad j = 1, \cdots, p, \tag{33}$$

$$f_{i,g-1}^{(g-1)} = -\frac{a_{i,g-1}^{(g-1)}}{a_{g-1,g-1}^{(g-1)}}, \qquad i = g, \cdots, p, \tag{34}$$

$$f_{ij}^{(g-1)} = 0, \qquad \text{其他}. \tag{35}$$

那么

$$a_{ij}^{(g)} = 0, \quad i = j+1, \cdots, p, \quad j = 1, \cdots, g-1, \tag{36}$$

$$a_{ij}^{(g)} = a_{ij}^{(g-1)}, \quad i = 1, \cdots, g-1, \quad j = 1, \cdots, p, \tag{37}$$

$$a_{ij}^{(g)} = a_{ij}^{(g-1)} + f_{i,g-1}^{(g-1)} a_{g-1,j}^{(g-1)} = a_{ij}^{(g-1)} - \frac{a_{i,g-1}^{(g-1)} a_{g-1,j}^{(g-1)}}{a_{g-1,g-1}^{(g-1)}}, \quad i, j = g, \cdots, p. \tag{38}$$

注意到 $\boldsymbol{F} = \boldsymbol{F}_{p-1}, \cdots, \boldsymbol{F}_1$ 是下三角矩阵, 而且 $\boldsymbol{A}_g$ 的前 $g-1$ 列对角线以下的元为 0, 显然,$\boldsymbol{A}^* = \boldsymbol{F}\boldsymbol{A}$ 是上三角矩阵. 由 $|\boldsymbol{A}| \neq 0$ 且 $|\boldsymbol{F}_{g-1}| = 1$, 我们有 $|\boldsymbol{A}_{g-1}| \neq 0$. 因此, $a_{11}^{(1)}, \cdots, a_{g-2,g-2}^{(g-2)}$ 都不为 0, 而且由 $\boldsymbol{A}_{g-1}$ 的最后 $p-g$ 列, 可以计算得 $a_{g-1,g-1}^{(g-1)} \neq 0$, 从而 $f_{i,g-1}^{(g-1)}$ 的定义有意义. ■

从等式 $\boldsymbol{FA}=\boldsymbol{A}^*$ 可以解得 $\boldsymbol{A}=\boldsymbol{LR}$, 其中 $\boldsymbol{R}=\boldsymbol{A}^*$ 是上三角矩阵, 而 $\boldsymbol{L}=\boldsymbol{F}^{-1}$ 是下三角矩阵, 且主对角线上的元素为 1(这是由于$\boldsymbol{F}$是下三角矩阵且主对角线上的元素为 1). 这就是著名的 LR分解.

推论 A.1.4 *如果 $\boldsymbol{A}$ 是正定的, 那么存在一个非奇异的下三角矩阵 $\boldsymbol{F}$, 使得$\boldsymbol{FAF'}$是对角矩阵, 并且是正定的.*

证明 由定理 A.1.2 可知, 存在一个非奇异的下三角矩阵 $\boldsymbol{F}$, 使得 $\boldsymbol{FA}$ 是一个非奇异的上三角矩阵. 从而 $\boldsymbol{FAF'}$ 是一个上三角矩阵, 并且是对称的, 因此它是对角矩阵. ■

推论 A.1.5 *正定矩阵 $\boldsymbol{A}$ 的行列式是正数.*

证明 从 $\boldsymbol{FAF'}$ 的构造可知,

$$\boldsymbol{FAF'}=\begin{pmatrix} a_{11}^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & a_{22}^{(2)} & 0 & \cdots & 0 \\ 0 & 0 & a_{33}^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{pp}^{(p)} \end{pmatrix} \tag{39}$$

是正定的, 因此 $a_{gg}^{(g)}>0, g=1,\cdots,p$, 而且 $0<|\boldsymbol{FAF'}|=|\boldsymbol{F}|\cdot|\boldsymbol{A}|\cdot|\boldsymbol{F}|=|\boldsymbol{A}|$. ■

推论 A.1.6 *如果 $\boldsymbol{A}$ 是正定的, 则存在下三角矩阵 $\boldsymbol{G}$ 使得 $\boldsymbol{GAG'}=\boldsymbol{I}$.*

证明 令 $\boldsymbol{FAF'}=\boldsymbol{D}^2$, 其中 $\boldsymbol{D}$ 为对角矩阵, 它的主对角线上的元素为 $\boldsymbol{D}^2$ 的主对角线上元素的算术平方根. 则由 $\boldsymbol{G}=\boldsymbol{D}^{-1}\boldsymbol{F}$ 就得结论成立. ■

推论 A.1.7 (Cholesky 分解) *如果 $\boldsymbol{A}$ 是正定的, 则存在唯一的一个主对角线上的元素都为正数的下三角矩阵$\boldsymbol{T}$ $(t_{ij}=0, i<j)$, 使得 $\boldsymbol{A}=\boldsymbol{TT'}$.*

证明 由推论 A.1.6, $\boldsymbol{A}=\boldsymbol{G}^{-1}(\boldsymbol{G}')^{-1}$, 其中$\boldsymbol{G}$ 是下三角矩阵. 则 $\boldsymbol{T}=\boldsymbol{G}^{-1}$ 是下三角矩阵. ■

事实上, 对于这个定理的 $\boldsymbol{A}=\boldsymbol{VV'}$ 的情形, 在 7.2 节已经得到了证明.

A.2 特征根和特征向量

方阵 $\boldsymbol{B}$ 的*特征根*定义为它的*特征方程*

$$|\boldsymbol{B}-\lambda\boldsymbol{I}|=0 \tag{1}$$

的根. 它或者为*隐式根*, 或者为*特征值*. 例如, 对于

$$\boldsymbol{B}=\begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix},$$

我们有

$$|\boldsymbol{B}-\lambda\boldsymbol{I}|=\begin{vmatrix} 5-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix}=25-4-10\lambda+\lambda^2=\lambda^2-10\lambda+21. \tag{2}$$

等式 (1) 中的多项式的次数为矩阵 $\boldsymbol{B}$ 的阶数, 而且常数项为 $|\boldsymbol{B}|$.

一个矩阵 $\boldsymbol{C}$ 称为是*正交的*, 如果 $\boldsymbol{C}'\boldsymbol{C}=\boldsymbol{I}$; 由这个定义立即可得 $\boldsymbol{CC}'=\boldsymbol{I}$. 令向量 $\boldsymbol{x}'=(x_1,\cdots,x_p)$ 和 $\boldsymbol{y}'=(y_1,\cdots,y_p)$ 代表 p 维欧氏空间中的两个点. 则它们之间距离的平

方为 $D(\boldsymbol{x},\boldsymbol{y})=(\boldsymbol{x}-\boldsymbol{y})'(\boldsymbol{x}-\boldsymbol{y})$. 变换 $\boldsymbol{z}=\boldsymbol{C}\boldsymbol{x}$ 可以看作是 p 维欧氏空间中的坐标轴的变换. 如果 $\boldsymbol{C}$ 是正交的, 那么这个变换是保持距离不变的, 因为

$$\begin{aligned}D(\boldsymbol{C}\boldsymbol{x},\boldsymbol{C}\boldsymbol{y})&=(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{C}\boldsymbol{x})'(\boldsymbol{C}\boldsymbol{y}-\boldsymbol{C}\boldsymbol{x})\\&=(\boldsymbol{y}-\boldsymbol{x})'\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{y}-\boldsymbol{x})=(\boldsymbol{y}-\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{x})=D(\boldsymbol{x},\boldsymbol{y}).\end{aligned}\tag{3}$$

因为一个三角形的角可以由它的边长所决定, 从而变换 $\boldsymbol{z}=\boldsymbol{C}\boldsymbol{x}$ 也保持角度不变. 正交变换是由一个或者多个坐标轴的旋转及可能的反射所组成的. 我们用 $\|\boldsymbol{x}\|$ 来表示 $\sqrt{\boldsymbol{x}'\boldsymbol{x}}$.

定理 A.2.1　*任意给定一个对称矩阵 $\boldsymbol{B}$, 存在一个正交矩阵 $\boldsymbol{C}$ 使得*

$$\boldsymbol{C}'\boldsymbol{B}\boldsymbol{C}=\boldsymbol{D}=\begin{pmatrix}d_1&0&\cdots&0\\0&d_2&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&d_p\end{pmatrix}.\tag{4}$$

如果 $\boldsymbol{B}$ 是半正定的, 那么 $d_i\geqslant 0, i=1,\cdots,p$; 如果 $\boldsymbol{B}$ 是正定的, 那么 $d_i>0, i=1,\cdots,p$.

在 11.2 节中, 讨论在 $\boldsymbol{B}$ 为半正定对称矩阵的情况下的主成分时已经给出这个定理的证明. 这样, 在$\boldsymbol{C}$的变换下, 特征方程 (1) 变成

$$\begin{aligned}0&=|\boldsymbol{C}'|\cdot|\boldsymbol{B}-\lambda\boldsymbol{I}|\cdot|\boldsymbol{C}|=|\boldsymbol{C}'(\boldsymbol{B}-\lambda\boldsymbol{I})\boldsymbol{C}|\\&=|\boldsymbol{C}'\boldsymbol{B}\boldsymbol{C}-\lambda\boldsymbol{I}|=|\boldsymbol{D}-\lambda\boldsymbol{I}|\\&=\begin{vmatrix}d_1-\lambda&0&\cdots&0\\0&d_2-\lambda&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&d_p-\lambda\end{vmatrix}=\prod_{i=1}^{p}(d_i-\lambda).\end{aligned}\tag{5}$$

从而,$\boldsymbol{B}$ 的特征根即是变换后的矩阵 $\boldsymbol{D}$ 的对角线上的元素.

如果 λ_i 是 $\boldsymbol{B}$ 的一个特征根, 那么一个满足

$$(\boldsymbol{B}-\lambda_i\boldsymbol{I})\boldsymbol{x}_i=\boldsymbol{0}\tag{6}$$

的非 $\boldsymbol{0}$ 向量 $\boldsymbol{x}_i$ 称为矩阵 $\boldsymbol{B}$ 关于特征根 λ_i 的**特征向量**. $\boldsymbol{x}_i$ 的任意数乘仍为一个特征向量. 当 $\boldsymbol{B}$ 是对称矩阵时, 有 $\boldsymbol{x}_i'(\boldsymbol{B}-\lambda_i\boldsymbol{I})=\boldsymbol{0}$. 如果特征根是互不相同的, 则有 $\boldsymbol{x}_j'\boldsymbol{B}\boldsymbol{x}_i=0$ 且 $\boldsymbol{x}_j'\boldsymbol{x}_i=0, i\neq j$. 设 $\boldsymbol{c}_i=(1/\|\boldsymbol{x}_i\|)\boldsymbol{x}_i$ 为第 i 个正规化的特征向量, 再设 $\boldsymbol{C}=(\boldsymbol{c}_1,\cdots,\boldsymbol{c}_p)$. 那么 $\boldsymbol{C}'\boldsymbol{C}=\boldsymbol{I}$, 而且 $\boldsymbol{B}\boldsymbol{C}=\boldsymbol{C}\boldsymbol{D}$. 这恰好符合 (4). 如果一个特征根是 m 重的, 那么对应于这个特征根的 m 个特征向量可以被它们的 m 个线性无关的线性组合所代替. 这样就可以选择一组向量满足 (6) 式, 以及 $\boldsymbol{x}_j'\boldsymbol{x}_i=0$ 和 $\boldsymbol{x}_j'\boldsymbol{B}\boldsymbol{x}_i=0,\ i\neq j$.

特征向量位于主轴的方向上 (参见第 11 章). 而且 $\boldsymbol{B}$ 的特征根与椭球面

$$\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}=1\tag{7}$$

的主轴长度倒数的平方成正比. 这是因为在旋转 $\boldsymbol{y}=\boldsymbol{C}\boldsymbol{x}$ 下, 有

$$1=\boldsymbol{y}'\boldsymbol{D}\boldsymbol{y}=\sum_{i=1}^{p}d_i{y_i}^2.\tag{8}$$

对于两个矩阵 $\boldsymbol{A}$(非奇异) 和 $\boldsymbol{B}$, 我们可以考虑它们之间的方程

$$|\boldsymbol{B}-\lambda\boldsymbol{A}|=0. \tag{9}$$

我们对这个方程的根感兴趣, 是因为它在一定的变换下具有不变性. 事实上, 对于任何一个非奇异的矩阵 $\boldsymbol{C}$, 方程

$$|\boldsymbol{C}'\boldsymbol{B}\boldsymbol{C}-\lambda\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}|=0 \tag{10}$$

的根与 (9) 相同, 这是因为

$$|\boldsymbol{C}'\boldsymbol{B}\boldsymbol{C}-\lambda\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}|=|\boldsymbol{C}'(\boldsymbol{B}-\lambda\boldsymbol{A})\boldsymbol{C}|=|\boldsymbol{C}'|\cdot|\boldsymbol{B}-\lambda\boldsymbol{A}|\cdot|\boldsymbol{C}| \tag{11}$$

而且, $|\boldsymbol{C}'|=|\boldsymbol{C}|\neq 0$.

由推论 A.1.6 可得, 如果 $\boldsymbol{A}$ 是正定的, 则存在一个矩阵 $\boldsymbol{E}$ 使得 $\boldsymbol{E}'\boldsymbol{A}\boldsymbol{E}=\boldsymbol{I}$. 令 $\boldsymbol{E}'\boldsymbol{B}\boldsymbol{E}=\boldsymbol{B}^*$. 从定理 A.2.1 中, 我们可以推导出存在一个正交矩阵 $\boldsymbol{C}$, 使得 $\boldsymbol{C}'\boldsymbol{B}^*C=\boldsymbol{D}$, 其中 $\boldsymbol{D}$ 为对角矩阵. 定义 $\boldsymbol{E}\boldsymbol{C}$ 为 $\boldsymbol{F}$. 我们有下面的定理.

定理 A.2.2 *任意给定矩阵 $\boldsymbol{B}$ 为半正定的, $\boldsymbol{A}$ 为正定的. 则存在一个非奇异矩阵 $\boldsymbol{F}$, 使得*

$$\boldsymbol{F}'\boldsymbol{B}\boldsymbol{F}=\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_p\end{pmatrix}, \tag{12}$$

$$\boldsymbol{F}'\boldsymbol{A}\boldsymbol{F}=\boldsymbol{I}, \tag{13}$$

其中 $\lambda_1\geqslant\cdots\geqslant\lambda_p\,(\geqslant 0)$ 是 (9) 的根. 如果 $\boldsymbol{B}$ 是正定的, 则 $\lambda_i>0, i=1,\cdots,p$.

对应于每一个根 λ_i 都有一个向量 $\boldsymbol{x}_i$, 满足

$$(\boldsymbol{B}-\lambda_i\boldsymbol{A})\boldsymbol{x}_i=\boldsymbol{0} \tag{14}$$

并且 $\boldsymbol{x}_i{}'\boldsymbol{A}\boldsymbol{x}_i=1$. 如果这些根互不相同, 则 $\boldsymbol{x}_j{}'\boldsymbol{B}\boldsymbol{x}_i=0$ 且 $\boldsymbol{x}_j{}'\boldsymbol{A}\boldsymbol{x}_i=0, i\neq j$. 这样, $\boldsymbol{F}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_p)$. 如果有一个根为 m 重的, 那么 m 个线性无关的向量 $\boldsymbol{x}_i$ 可以被它们的 m 个线性无关的线性组合所代替. 这样就可以选择一组向量, 使得它们满足 (14) 式, 并且满足 $\boldsymbol{x}_j{}'\boldsymbol{B}\boldsymbol{x}_i=0$ 和 $\boldsymbol{x}_j{}'\boldsymbol{A}\boldsymbol{x}_i=0, i\neq j$.

定理 A.2.3 (奇异值分解) *给定一个 $n\times p$ 矩阵 $\boldsymbol{X}$, $n\geqslant p$, 则存在一个 $n\times n$ 正交矩阵 $\boldsymbol{P}$, 一个 $p\times p$ 正交矩阵 $\boldsymbol{Q}$, 以及一个由 $p\times p$ 的半正定对角矩阵和 $(n-p)\times p$ 的零矩阵组成的 $n\times p$ 矩阵 $\boldsymbol{D}$, 满足*

$$\boldsymbol{X}=\boldsymbol{P}\boldsymbol{D}\boldsymbol{Q}. \tag{15}$$

证明 由定理 A.2.1, 存在一个 $p\times p$ 的正交矩阵$\boldsymbol{Q}$和一个对角矩阵 $\boldsymbol{E}_1$, 使得

$$\boldsymbol{Q}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q}'=\begin{pmatrix}\boldsymbol{E}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}, \tag{16}$$

其中 $\boldsymbol{E}_1$ 是正定的对角矩阵. 令 $\boldsymbol{X}\boldsymbol{Q}'=\boldsymbol{Y}=(\boldsymbol{Y}_1\ \boldsymbol{Y}_2)$, $\boldsymbol{Y}_1$ 的列数与 $\boldsymbol{E}_1$ 的阶相同. 从而 $\boldsymbol{Y}_2{}'\boldsymbol{Y}_2=\boldsymbol{0}$, 因此 $\boldsymbol{Y}_2=\boldsymbol{0}$. 令 $\boldsymbol{P}_1=\boldsymbol{Y}_1\boldsymbol{E}_1{}^{-\frac{1}{2}}$, 则 $\boldsymbol{P}_1{}'\boldsymbol{P}_1=\boldsymbol{I}$. 通过添加一个矩阵 $\boldsymbol{P}_2$ 使得 $\boldsymbol{P}=(\boldsymbol{P}_1\ \boldsymbol{P}_2)$ 成为一个正交矩阵, 就得到了满足定理的一个 $n\times n$ 的正交矩阵 $\boldsymbol{P}$. 矩阵 $\boldsymbol{D}$ 的左上角一块为 $\boldsymbol{E}_1{}^{\frac{1}{2}}$, 其他部分都由 0 组成. ∎

定理 A.2.4 *设 $\boldsymbol{A}$ 为正定矩阵,$\boldsymbol{B}$ 为半正定矩阵. 那么有*

$$\lambda_p\leqslant\frac{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}}\leqslant\lambda_1, \tag{17}$$

其中 λ_1 和 λ_p 分别为 (1) 式的最大的根和最小的根, 同样有

$$\lambda_p\leqslant\frac{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}\leqslant\lambda_1, \tag{18}$$

其中 λ_1 和 λ_p 分别为 (9) 式的最大的根和最小的根.

证明 不等式 (17) 本质上已经在 11.2 节得到了证明, 也可以由 (4) 推得. 不等式 (18) 是定理 A.2.2 的直接结果. ■

方阵 $\boldsymbol{A}$ 称为**幂等的**, 如果 $\boldsymbol{A}^2=\boldsymbol{A}$. 如果 λ 满足 $|\boldsymbol{A}-\lambda\boldsymbol{I}|=0$, 则存在一个向量 $\boldsymbol{x}\neq\boldsymbol{0}$, 使得 $\lambda\boldsymbol{x}=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^2\boldsymbol{x}$. 然而, $\boldsymbol{A}^2\boldsymbol{x}=\boldsymbol{A}(\boldsymbol{A}\boldsymbol{x})=\boldsymbol{A}\lambda\boldsymbol{x}=\lambda^2\boldsymbol{x}$. 因此 $\lambda^2=\lambda$, 故 λ 等于 0 或者 1. $\lambda=1$ 的重数是$\boldsymbol{A}$的秩. 如果$\boldsymbol{A}$是 $p\times p$ 矩阵, 那么 $\boldsymbol{I}_p-A$ 是秩为 $p-(\text{rank }\boldsymbol{A})$ 的幂等矩阵, 并且 $\boldsymbol{A}$ 与 $\boldsymbol{I}_p-A$ 相互正交. 如果$\boldsymbol{A}$是对称的, 则存在正交矩阵 $\boldsymbol{O}$, 使得

$$\boldsymbol{O}\boldsymbol{A}\boldsymbol{O}'=\begin{bmatrix}\boldsymbol{I}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{0}\end{bmatrix},\quad \boldsymbol{O}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{O}'=\begin{bmatrix}\boldsymbol{0}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{I}\end{bmatrix}. \tag{19}$$

A.3 分块向量和分块矩阵

考虑 A.1 节中由 (1) 式所定义的矩阵 $\boldsymbol{A}$. 令

$$\begin{aligned}\boldsymbol{A}_{11}&=(a_{ij}),\quad i=1,\cdots,p,\quad j=1,\cdots,q,\\\boldsymbol{A}_{12}&=(a_{ij}),\quad i=1,\cdots,p,\quad j=q+1,\cdots,n,\\\boldsymbol{A}_{21}&=(a_{ij}),\quad i=p+1,\cdots,m,\quad j=1,\cdots,q,\\\boldsymbol{A}_{22}&=(a_{ij}),\quad i=p+1,\cdots,m,\quad j=q+1,\cdots,n,\end{aligned} \tag{1}$$

那么, 我们可以写成

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\\boldsymbol{A}_{21}&\boldsymbol{A}_{22}\end{pmatrix}. \tag{2}$$

我们称 $\boldsymbol{A}$ 被分块成子矩阵 $\boldsymbol{A}_{ij}$. 将 $\boldsymbol{B}$ $(m\times n)$ 按同样的方式分块成子矩阵 $\boldsymbol{B}_{ij}$, $i,j=1,2$. 那么

$$\boldsymbol{A}+\boldsymbol{B}=\begin{pmatrix}\boldsymbol{A}_{11}+\boldsymbol{B}_{11}&\boldsymbol{A}_{12}+\boldsymbol{B}_{12}\\\boldsymbol{A}_{21}+\boldsymbol{B}_{21}&\boldsymbol{A}_{22}+\boldsymbol{B}_{22}\end{pmatrix}, \tag{3}$$

现在将$\boldsymbol{C}$ $(n\times r)$ 进行如下的分块

$$\boldsymbol{C}=\begin{pmatrix}\boldsymbol{C}_{11}&\boldsymbol{C}_{12}\\\boldsymbol{C}_{21}&\boldsymbol{C}_{22}\end{pmatrix}, \tag{4}$$

其中 $\boldsymbol{C}_{11}$ 和 $\boldsymbol{C}_{12}$ 有 q 行, 而 $\boldsymbol{C}_{11}$ 和 $\boldsymbol{C}_{21}$ 有 s 列. 那么

$$\begin{aligned}\boldsymbol{AC}&=\begin{pmatrix}\boldsymbol{A}_{11}&\boldsymbol{A}_{12}\\\boldsymbol{A}_{21}&\boldsymbol{A}_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{C}_{11}&\boldsymbol{C}_{12}\\\boldsymbol{C}_{21}&\boldsymbol{C}_{22}\end{pmatrix}\\&=\begin{pmatrix}\boldsymbol{A}_{11}\boldsymbol{C}_{11}+\boldsymbol{A}_{12}\boldsymbol{C}_{21}&\boldsymbol{A}_{11}\boldsymbol{C}_{12}+\boldsymbol{A}_{12}\boldsymbol{C}_{22}\\\boldsymbol{A}_{21}\boldsymbol{C}_{11}+\boldsymbol{A}_{22}\boldsymbol{C}_{21}&\boldsymbol{A}_{21}\boldsymbol{C}_{12}+\boldsymbol{A}_{22}\boldsymbol{C}_{22}\end{pmatrix}.\end{aligned} \tag{5}$$

为了验证这一点, 首先考虑 $\boldsymbol{AC}$ 的前 p 行和前 s 列的元素. 第 (i,j) 元是

$$\sum_{k=1}^{n}a_{ik}c_{kj},\quad i\leqslant p,\quad j\leqslant s. \tag{6}$$

这个和式可以写成

$$\sum_{k=1}^{q}a_{ik}c_{kj}+\sum_{k=q+1}^{n}a_{ik}c_{kj}. \tag{7}$$

第一个和式是 $A_{11}C_{11}$ 的第 (i,j) 元, 第二个和式是 $A_{12}C_{21}$ 的第 (i,j) 元, 因此, 总的和式 (6) 是 $A_{11}C_{11}+A_{12}C_{21}$ 的 (i,j) 元. 用同样的方法, 我们可以验证 AC 的其他子矩阵也可以写成 (5) 的形式.

我们顺便提及一下, 如果 A 按照 (2) 进行分块, 那么 A 的转置可以写成

$$A' = \begin{pmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{pmatrix}. \tag{8}$$

如果 $A_{12}=0$ 且 $A_{21}=0$, 那么对于正定矩阵 A 以及方阵 A_{11}, 我们有

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix}. \tag{9}$$

等号右边的矩阵是存在的, 因为 A_{11} 和 A_{22} 是非奇异的. 右边的矩阵确实为 A 的逆矩阵, 这一点可以用乘法来验证:

$$\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \tag{10}$$

它是矩阵 I_p 的分块形式.

我们还要指出

$$\begin{vmatrix} A_{11} & 0 \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & 0 \\ 0 & I \end{vmatrix} \cdot \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix} = |A_{11}|\cdot|A_{22}|. \tag{11}$$

中间部分的第一个行列式, 可以根据最后一行元素的子式的展开得到; 展开后的和式中, 只有最后一项是唯一的非零项, 它等于 1 乘以一个行列式, 这个行列式与之前的行列式有同样的形式, 只是 I 的阶数少 1. 重复这个过程, 直到 $|A_{11}|$ 就是子式. 同样,

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix} \cdot \begin{vmatrix} A_{11} & A_{12} \\ 0 & I \end{vmatrix} = |A_{11}|\cdot|A_{22}|. \tag{12}$$

一个有用的事实是, 如果 A_1 是一个 q 行 p 列的秩为 q 的矩阵, 那么存在一个 $p-q$ 行 p 列的矩阵 A_2 使得

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \tag{13}$$

是非奇异的. 这个结论可以用以下的方式进行验证. 计算 A 的列, 使得由 A_1 的前 q 列组成的矩阵 A_{11} 是非奇异的 (A_1 至少有一个 $q\times q$ 子式非零), 然后记 A_2 为 $(0\ I)$, 这样

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ 0 & I \end{vmatrix} = |A_{11}|, \tag{14}$$

而它不等于零.

定理 A.3.1 *将方阵 A 按照 (2) 进行分块, 使得 A_{22} 为方阵. 如果 A_{22} 是非奇异的, 设*

$$B = \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}, \quad C = \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix}. \tag{15}$$

那么

$$BA = \begin{pmatrix} A_{11}-A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \quad AC = \begin{pmatrix} A_{11}-A_{12}A_{22}^{-1}A_{21} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \tag{16}$$

$$BAC = \begin{pmatrix} A_{11}-A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{pmatrix}. \tag{17}$$

如果 $\boldsymbol{A}$ 是对称的, 则 $\boldsymbol{C}=\boldsymbol{B}'$.

定理 A.3.2 将方阵 $\boldsymbol{A}$ 按照 (2) 进行分块, 使得 $\boldsymbol{A}_{22}$ 为方阵. 如果 $\boldsymbol{A}_{22}$ 是非奇异的, 则

$$|\boldsymbol{A}|=\left|\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right|\cdot\left|\boldsymbol{A}_{22}\right|. \tag{18}$$

证明 因为在 (16) 中有 $|\boldsymbol{B}|=1$, 从而可得 (18) 式. ■

推论 A.3.1 对于非奇异矩阵 $\boldsymbol{C}$, 有

$$\begin{vmatrix}\boldsymbol{C} & \boldsymbol{y}\\ \boldsymbol{y}' & 1\end{vmatrix}=|\boldsymbol{C}-\boldsymbol{y}\boldsymbol{y}'|=\begin{vmatrix}1 & \boldsymbol{y}'\\ \boldsymbol{y} & \boldsymbol{C}\end{vmatrix}=|\boldsymbol{C}|(1-\boldsymbol{y}'\boldsymbol{C}^{-1}\boldsymbol{y}). \tag{19}$$

定理 A.3.3 将非奇异矩阵 $\boldsymbol{A}$ 按照 (2) 进行分块, 使得 $\boldsymbol{A}_{22}$ 为方阵. 如果 $\boldsymbol{A}_{22}$ 是非奇异的, 令 $\boldsymbol{A}_{11\cdot2}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$. 那么

$$\boldsymbol{A}^{-1}=\begin{pmatrix}\boldsymbol{A}_{11\cdot2}^{-1} & -\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot2}^{-1} & \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\end{pmatrix}. \tag{20}$$

证明 由定理 A.3.1,

$$\boldsymbol{A}=\boldsymbol{B}^{-1}\begin{pmatrix}\boldsymbol{A}_{11\cdot2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{22}\end{pmatrix}\boldsymbol{C}^{-1}. \tag{21}$$

因此,

$$\begin{aligned}\boldsymbol{A}^{-1}&=\boldsymbol{C}\begin{pmatrix}\boldsymbol{A}_{11\cdot2} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{22}\end{pmatrix}^{-1}\boldsymbol{B} \\ &=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{A}_{11\cdot2}^{-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{22}^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{I} & -\boldsymbol{A}_{12}\boldsymbol{A}_{22}\\ \boldsymbol{0} & \boldsymbol{I}\end{pmatrix}.\end{aligned} \tag{22}$$

直接进行乘法运算, 就可以得到想要的结果. ■

推论 A.3.2 如果 $\boldsymbol{x}'=(\boldsymbol{x}^{(1)\prime}\ \boldsymbol{x}^{(2)\prime})$, 那么,

$$\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}=\left(\boldsymbol{x}^{(1)}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{x}^{(2)}\right)'\boldsymbol{A}_{11\cdot2}^{-1}\left(\boldsymbol{x}^{(1)}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{x}^{(2)}\right)+\boldsymbol{x}^{(2)\prime}\boldsymbol{A}_{22}^{-1}\boldsymbol{x}^{(2)}. \tag{23}$$

证明 根据定理, 有

$$\begin{aligned}\boldsymbol{x}'\boldsymbol{A}^{-1}\boldsymbol{x}=&\,\boldsymbol{x}^{(1)\prime}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{x}^{(1)}-\boldsymbol{x}^{(1)\prime}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{x}^{(2)} \\ &-\boldsymbol{x}^{(2)\prime}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{x}^{(1)}+\boldsymbol{x}^{(2)\prime}\left(\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\right)\boldsymbol{x}^{(2)},\end{aligned} \tag{24}$$

它恰好等于 (23) 的右边. ■

定理 A.3.4 将非奇异矩阵 $\boldsymbol{A}$ 按照 (2) 进行分块, 使得 $\boldsymbol{A}_{22}$ 为方阵. 如果 $\boldsymbol{A}_{22}$ 是非奇异的, 则

$$\left(\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\right)^{-1}=\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\left(\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\right)^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}. \tag{25}$$

证明 由定理 A.3.3 可知 $\boldsymbol{A}^{-1}$ 的右下方的子块恰等于 (25) 的右边, 而且交换 1 和 2 后, 它也等于 (25) 的左边. ■

定理 A.3.5 设 $\boldsymbol{U}$ 是一个 $p\times m$ 矩阵. 则 $\boldsymbol{I}_p-\boldsymbol{U}\boldsymbol{U}'$, $\boldsymbol{I}_m-\boldsymbol{U}'\boldsymbol{U}$ 以及

$$\begin{pmatrix}\boldsymbol{I}_p & \boldsymbol{U}\\ \boldsymbol{U}' & \boldsymbol{I}_m\end{pmatrix} \tag{26}$$

为正定矩阵的条件相同.

证明 我们有

$$(v'\ w')\begin{pmatrix} I_p & U \\ U' & I_m \end{pmatrix}\begin{pmatrix} v \\ w \end{pmatrix} = v'v + v'Uw + w'U'v + w'w \tag{27}$$
$$= v'(I_m - UU')v + (U'v + w)'(U'v + w).$$

等式右边的第二项是非负的; 对于所有的 $v \neq 0$ 第一项是正数, 当且仅当 $I_m - U'U$ 是正定的. 交换 v 和 w 的位置, 则 (26) 是正定的, 当且仅当 $I_p - UU'$ 是正定的. ■

A.4 其他方面的一些结果

定理 A.4.1 设 C 是 $p \times p$ 半正定矩阵, 并且秩为 r $(r \leqslant p)$. 那么存在一个非奇异矩阵 A, 使得

$$ACA' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{1}$$

证明 因为 C 的秩为 r, 存在一个 $(p-r) \times p$ 矩阵 A_2, 使得

$$A_2 C = 0. \tag{2}$$

选择B $(r \times p)$, 使得

$$\begin{pmatrix} B \\ A_2 \end{pmatrix} \tag{3}$$

是非奇异的. 那么,

$$\begin{pmatrix} B \\ A_2 \end{pmatrix} C(B'\ A_2') = \begin{pmatrix} BC \\ 0 \end{pmatrix}(B'\ A_2') = \begin{pmatrix} BCB' & 0 \\ 0 & 0 \end{pmatrix}. \tag{4}$$

这个矩阵的秩为 r, 因此 BCB' 是非奇异的. 由推论 A.1.6 知, 存在一个非奇异矩阵 D, 使得 $D(BCB')D' = I_r$. 那么

$$A = \begin{pmatrix} DB \\ A_2 \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} B \\ A_2 \end{pmatrix} \tag{5}$$

是一个非奇异矩阵, 使得 (1) 成立. ■

引理 A.4.1 如果 E 是 $p \times p$ 的对称非奇异矩阵, 那么存在一个非奇异矩阵 F, 使得

$$FEF' = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{6}$$

其中 I 的阶数等于 E 的正特征根的个数, 而 $-I$ 的阶数等于 E 的负特征根的个数.

证明 由定理 A.2.1 我们知道, 存在一个正交矩阵 G, 使得

$$GEG' = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & h_p \end{pmatrix}, \tag{7}$$

其中 $h_1 \geqslant \cdots \geqslant h_q > 0 > h_{q+1} \geqslant \cdots \geqslant h_p$ 是 E 的特征根. 令

$$
\boldsymbol{K}=\begin{pmatrix}
1/\sqrt{h_1} & \cdots & 0 & 0 & \cdots & 0\\
\vdots & & \vdots & \vdots & & \vdots\\
0 & \cdots & 1/\sqrt{h_q} & 0 & \cdots & 0\\
0 & \cdots & 0 & 1/\sqrt{-h_{q+1}} & \cdots & 0\\
\vdots & & \vdots & \vdots & & \vdots\\
0 & \cdots & 0 & 0 & \cdots & 1/\sqrt{-h_p}
\end{pmatrix}. \tag{8}
$$

那么

$$
\boldsymbol{KGEG'K'}=(\boldsymbol{KG})\boldsymbol{E}(\boldsymbol{KG})'=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{I}\end{pmatrix}. \quad ■ \tag{9}
$$

推论 A.4.1　设 $\boldsymbol{C}$ 是 $p\times p$ 的对称矩阵, 秩为 r $(\leqslant p)$. 那么存在一个非奇异矩阵 $\boldsymbol{A}$, 使得

$$
\boldsymbol{ACA'}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}, \tag{10}
$$

其中$\boldsymbol{I}$的阶数等于 $\boldsymbol{C}$ 的正特征根的个数, 而 $-\boldsymbol{I}$ 的阶数等于 $\boldsymbol{C}$ 的负特征根的个数, 并且阶数的和等于 r.

证明　除了用引理 A.4.1 代替推论 A.1.6 之外, 这个证明与定理 A.4.1 的证明相同. ■

引理 A.4.2　设 $\boldsymbol{A}$ 是 $n\times m$ 矩阵 $(n>m)$, 满足

$$
\boldsymbol{A'A}=\boldsymbol{I}_m. \tag{11}
$$

存在一个 $n\times(n-m)$ 矩阵 $\boldsymbol{B}$, 使得 $(\boldsymbol{AB})$ 为正交矩阵.

证明　因为$\boldsymbol{A}$的秩为 m, 存在一个 $n\times(n-m)$ 矩阵 $\boldsymbol{C}$, 使得 $(\boldsymbol{AC})$ 为非奇异的. 记 $\boldsymbol{C}-\boldsymbol{AA'C}$ 为$\boldsymbol{D}$, 那么 $\boldsymbol{D'A}=\boldsymbol{0}$. 设 $\boldsymbol{E}$ $[(n-m)\times(n-m)]$ 满足 $\boldsymbol{E'D'DE}=\boldsymbol{I}$. 那么, 记$\boldsymbol{DE}$为 $\boldsymbol{B}$, 即为所求. ■

引理 A.4.3　设 $\boldsymbol{x}$ 是一个 n 维向量, 那么存在一个正交矩阵 $\boldsymbol{O}$, 使得

$$
\boldsymbol{Ox}=\begin{pmatrix}c\\0\\\vdots\\0\end{pmatrix}, \tag{12}
$$

其中 $c=\sqrt{\boldsymbol{x'x}}$.

证明　令 $\boldsymbol{O}$ 的第一行为 $(1/c)\ \boldsymbol{x}'$. 然后, 可以按任何一种方式选择其他行, 使其成为一个正交矩阵. ■

引理 A.4.4　设 $\boldsymbol{B}=(b_{ij})$ 是一个 $p\times p$ 矩阵. 那么

$$
\frac{\partial|\boldsymbol{B}|}{\partial b_{ij}}=B_{ij},\quad i,j=1,\cdots,p. \tag{13}
$$

证明　将 $|\boldsymbol{B}|$ 按第 i 行的元素展开, 有

$$
|\boldsymbol{B}|=\sum_{h=1}^{p}b_{ih}B_{ih}. \tag{14}
$$

因为 B_{ih} 中不包含 b_{ij}, 引理得证. ■

引理 A.4.5 设 $b_{ij}=\beta_{ij}(c_1,\cdots,c_n)$ 是 $p\times p$ 矩阵 $\boldsymbol{B}$ 的第 (i,j) 元. 那么, 对于 $g=1,\cdots,n$, 有

$$\frac{\partial|\boldsymbol{B}|}{\partial c_g}=\sum_{i,h=1}^{p}\frac{\partial|\boldsymbol{B}|}{\partial b_{ih}}\cdot\frac{\partial\beta_{ih}(c_1,\cdots,c_n)}{\partial c_g}=\sum_{i,h=1}^{p}B_{ih}\frac{\partial\beta_{ih}(c_1,\cdots,c_n)}{\partial c_g}. \tag{15}$$

定理 A.4.2 如果 $\boldsymbol{A}=\boldsymbol{A}'$, 那么

$$\frac{\partial|\boldsymbol{A}|}{\partial a_{ii}}=A_{ii}, \tag{16}$$

$$\frac{\partial|\boldsymbol{A}|}{\partial a_{ij}}=2A_{ij},\quad i\neq j. \tag{17}$$

证明 将 $|\boldsymbol{A}|$ 按照第 i 行的元素展开, 即可得到 (16). 为了证明 (17), 令 $b_{ij}=b_{ji}=a_{ij}$, $i,j=1,\cdots,p$, $i\leqslant j$. 那么由引理 A.4.5,

$$\frac{\partial|\boldsymbol{B}|}{\partial a_{ij}}=B_{ij}+B_{ji}, \tag{18}$$

由于 $|\boldsymbol{A}|=|\boldsymbol{B}|$, 而且 $B_{ij}=B_{ji}=A_{ij}=A_{ji}$. 从而 (17) 得证. ■

定理 A.4.3

$$\frac{\partial}{\partial\boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})=2\boldsymbol{A}\boldsymbol{x}, \tag{19}$$

其中 $\partial/\partial\boldsymbol{x}$ 表示对$\boldsymbol{x}$的每一个分量求偏导数, 并且将所得到的偏导数排成一列.

证明 设 $\boldsymbol{h}$ 是与 $\boldsymbol{x}$ 分量相同的列向量. 那么,

$$\begin{aligned}(\boldsymbol{x}+\boldsymbol{h})'\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{h})&=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}+\boldsymbol{h}'\boldsymbol{A}\boldsymbol{x}+\boldsymbol{x}'\boldsymbol{A}\boldsymbol{h}+\boldsymbol{h}'\boldsymbol{A}\boldsymbol{h}\\&=\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}+2\boldsymbol{h}'\boldsymbol{A}\boldsymbol{x}+\boldsymbol{h}'\boldsymbol{A}\boldsymbol{h}.\end{aligned} \tag{20}$$

偏导数向量就是在第二项中右乘 $\boldsymbol{h}'$ 的向量. ■

定义 A.4.1 设 $\boldsymbol{A}=(a_{ij})$ 是一个 $p\times m$ 矩阵,$\boldsymbol{B}=(b_{\alpha\beta})$ 是一个 $q\times n$ 矩阵. 则第 i,α 行第 j,β 列的元素为 $a_{ij}b_{\alpha\beta}$ 的 $pq\times mn$ 矩阵称为矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的**Kronecker 乘积或者直积**, 记作 $\boldsymbol{A}\otimes\boldsymbol{B}$; 也就是,

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{pmatrix}a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1m}\boldsymbol{B}\\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2m}\boldsymbol{B}\\ \vdots & \vdots & & \vdots\\ a_{p1}\boldsymbol{B} & a_{p2}\boldsymbol{B} & \cdots & a_{pm}\boldsymbol{B}\end{pmatrix}. \tag{21}$$

当这些矩阵的阶允许它们进行相应的运算时, 有下面的一些性质成立.

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})=(\boldsymbol{A}\boldsymbol{C})\otimes(\boldsymbol{B}\boldsymbol{D}), \tag{22}$$

$$(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}. \tag{23}$$

定理 A.4.4 设 $\boldsymbol{A}$ $(p\times p)$ 的第 i 个特征根为 λ_i, 相应的特征向量为 $\boldsymbol{x}_i=(x_{1i},\cdots,x_{pi})'$, 而 $\boldsymbol{B}$ $(q\times q)$ 的第 α 个特征根为 ν_α, 相应的特征向量为 $\boldsymbol{y}_\alpha,\alpha=1,\cdots,q$. 那么 $\boldsymbol{A}\otimes\boldsymbol{B}$ 的第 i,α 个特征根为 $\lambda_i\nu_\alpha$, 并且相对应的特征向量为 $\boldsymbol{x}_i\otimes\boldsymbol{y}_\alpha=(x_{1i}\boldsymbol{y}'_\alpha,\cdots,x_{pi}\boldsymbol{y}'_\alpha)'$, $i=1,\cdots,p$, $\alpha=1,\cdots,q$.

证明

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{x}_i\otimes\boldsymbol{y}_\alpha)=\begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1p}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{p1}\boldsymbol{B} & \cdots & a_{pp}\boldsymbol{B}\end{pmatrix}\begin{pmatrix} x_{1i}\boldsymbol{y}_\alpha \\ \vdots \\ x_{pi}\boldsymbol{y}_\alpha\end{pmatrix} \tag{24}$$

$$=\begin{pmatrix} \sum_j a_{1j}x_{ji}\boldsymbol{B}\boldsymbol{y}_\alpha \\ \vdots \\ \sum_j a_{pj}x_{ji}\boldsymbol{B}\boldsymbol{y}_\alpha \end{pmatrix}$$

$$=\begin{pmatrix} \lambda_i x_{1i}\boldsymbol{B}\boldsymbol{y}_\alpha \\ \vdots \\ \lambda_i x_{pi}\boldsymbol{B}\boldsymbol{y}_\alpha \end{pmatrix}=\lambda_i\nu_\alpha\begin{pmatrix} x_{1i}\boldsymbol{y}_\alpha \\ \vdots \\ x_{pi}\boldsymbol{y}_\alpha\end{pmatrix}. \qquad \blacksquare$$

定理 A.4.5

$$|\boldsymbol{A}\otimes\boldsymbol{B}|=|\boldsymbol{A}|^q\,|\boldsymbol{B}|^p. \tag{25}$$

证明 任何一个矩阵的行列式等于它的所有的特征根的乘积, 因此

$$|\boldsymbol{A}\otimes\boldsymbol{B}|=\prod_{i=1}^p\prod_{\alpha=1}^q\lambda_i\nu_\alpha=\left(\prod_{i=1}^p\lambda_i\right)^q\left(\prod_{\alpha=1}^q\nu_\alpha\right)^p. \qquad \blacksquare \tag{26}$$

定义 A.4.2 如果 $p\times m$ 矩阵 $\boldsymbol{A}=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m)$, 那么它的拉直向量 $\mathrm{vec}\,\boldsymbol{A}=(\boldsymbol{a}_1',\cdots,\boldsymbol{a}_m')'$.

关于拉直向量的一些性质如下 [例如, Magnus(1988)]

$$\mathrm{vec}\,\boldsymbol{ABC}=(\boldsymbol{C}'\otimes\boldsymbol{A})\,\mathrm{vec}\,\boldsymbol{B}, \tag{27}$$

$$\mathrm{vec}\,\boldsymbol{x}\boldsymbol{y}'=\boldsymbol{y}\otimes\boldsymbol{x}. \tag{28}$$

定理 A.4.6 变换 $\boldsymbol{E}=\boldsymbol{Y}^{-1}$(从 $\boldsymbol{E}$ 到 $\boldsymbol{Y}$) 的雅可比行列式为 $|\boldsymbol{Y}|^{-2p}$, 其中 p 为$\boldsymbol{E}$和$\boldsymbol{Y}$的阶数.

证明 由 $\boldsymbol{EY}=\boldsymbol{I}$, 我们有

$$\left(\frac{\partial}{\partial\theta}\boldsymbol{E}\right)\boldsymbol{Y}+\boldsymbol{E}\left(\frac{\partial}{\partial\theta}\boldsymbol{Y}\right)=\boldsymbol{0}, \tag{29}$$

其中

$$\left(\frac{\partial}{\partial\theta}\boldsymbol{E}\right)=\begin{pmatrix} \frac{\partial e_{11}}{\partial\theta} & \cdots & \frac{\partial e_{1p}}{\partial\theta} \\ \vdots & & \vdots \\ \frac{\partial e_{p1}}{\partial\theta} & \cdots & \frac{\partial e_{pp}}{\partial\theta}\end{pmatrix}. \tag{30}$$

这样

$$\left(\frac{\partial}{\partial\theta}\boldsymbol{E}\right)=-\boldsymbol{E}\left(\frac{\partial}{\partial\theta}\boldsymbol{Y}\right)\boldsymbol{E}=-\boldsymbol{Y}^{-1}\left(\frac{\partial}{\partial\theta}\boldsymbol{Y}\right)\boldsymbol{Y}^{-1}. \tag{31}$$

如果 $\theta=y_{\alpha\beta}$, 那么

$$\left(\frac{\partial}{\partial y_{\alpha\beta}}\boldsymbol{E}\right)=-\boldsymbol{E}\boldsymbol{\varepsilon}_{\alpha\beta}\boldsymbol{E}=-\,\boldsymbol{e}_{\cdot\alpha}\,\boldsymbol{e}_{\beta\cdot}\,, \tag{32}$$

其中 $\boldsymbol{\varepsilon}_{\alpha\beta}$ 是一个 $p\times p$ 矩阵, 它除了第 (α,β) 元素为 1 外, 其他的元素都为 0. 而且 $\boldsymbol{e}_{\cdot\alpha}$ 是 $\boldsymbol{E}$ 的第 α 列,$\boldsymbol{e}_{\beta\cdot}$ 是它的第 β 行. 因此 $\partial e_{ij}/\partial y_{\alpha\beta}=-e_{i\alpha}e_{\beta j}$. 这样, 雅可比行列式就是一个 $p^2\times p^2$ 矩阵的行列式

$$\operatorname{mod}\left|\frac{\partial e_{ij}}{\partial y_{\alpha\beta}}\right| = |e_{i\alpha}e_{\beta j}| = |\boldsymbol{E}\otimes\boldsymbol{E}'| = |\boldsymbol{E}|^p|\boldsymbol{E}'|^p = |\boldsymbol{E}|^{2p} = |\boldsymbol{Y}|^{-2p}. \quad\blacksquare \tag{33}$$

定理 A.4.7 设 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是对称矩阵, 它们的特征根分别为 $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_p$ 和 $b_1 \geqslant b_2 \geqslant \cdots \geqslant b_p$, 设 $\boldsymbol{H}$ 是一个 $p\times p$ 正交矩阵. 则

$$\max_{\boldsymbol{H}} \operatorname{tr} \boldsymbol{HAH'B} = \sum_{j=1}^{p} a_j b_j, \qquad \min_{\boldsymbol{H}} \operatorname{tr} \boldsymbol{HAH'B} = \sum_{j=1}^{p} a_j b_{p+1-j}. \tag{34}$$

证明 设 $\boldsymbol{A} = \boldsymbol{H}_a\boldsymbol{D}_a\boldsymbol{H}_a'$, $\boldsymbol{B} = \boldsymbol{H}_b\boldsymbol{D}_b\boldsymbol{H}_b'$, 其中 $\boldsymbol{H}_a$ 和 $\boldsymbol{H}_b$ 是正交矩阵, 并且 $\boldsymbol{D}_a$ 和 $\boldsymbol{D}_b$ 是主对角线上元素分别为 $a_1,\cdots,a_p$ 和 $b_1,\cdots,b_p$ 的对角矩阵. 那么

$$\begin{aligned}\max_{\boldsymbol{H}^*} \operatorname{tr} \boldsymbol{H}^*\boldsymbol{A}\boldsymbol{H}^{*\prime}\boldsymbol{B} &= \max_{\boldsymbol{H}^*} \operatorname{tr} \boldsymbol{H}^*\boldsymbol{H}_a\boldsymbol{D}_a\boldsymbol{H}_a'\boldsymbol{H}^{*\prime}\boldsymbol{H}_b\boldsymbol{D}_b\boldsymbol{H}_b' \\ &= \max_{\boldsymbol{H}^*} \operatorname{tr} \boldsymbol{H}_b'\boldsymbol{H}^*\boldsymbol{H}_a\boldsymbol{D}_a(\boldsymbol{H}_b'\boldsymbol{H}^*\boldsymbol{H}_a)'\boldsymbol{D}_b \\ &= \max_{\boldsymbol{H}} \operatorname{tr} \boldsymbol{H}\boldsymbol{D}_a\boldsymbol{H}'\boldsymbol{D}_b,\end{aligned} \tag{35}$$

其中,$\boldsymbol{H} = \boldsymbol{H}_b'\boldsymbol{H}^*\boldsymbol{H}_a$. 根据下面的引理 A.4.6, 我们有

$$\begin{aligned}\operatorname{tr} \boldsymbol{H}\boldsymbol{D}_a\boldsymbol{H}'\boldsymbol{D}_b &= \sum_{i=1}^{p}(\boldsymbol{H}\boldsymbol{D}_a\boldsymbol{H}')_{ii}b_i \\ &= \sum_{i=1}^{p-1}\sum_{j=1}^{i}(\boldsymbol{H}\boldsymbol{D}_a\boldsymbol{H}')_{jj}(b_i - b_{i+1}) + b_p\sum_{j=1}^{p}(\boldsymbol{H}\boldsymbol{D}_a\boldsymbol{H}')_{jj} \\ &\leqslant \sum_{i=1}^{p-1}\sum_{j=1}^{i} a_j(b_i - b_{i+1}) + b_p\sum_{j=1}^{p} a_j \\ &= \sum_{i=1}^{p} a_i b_i.\end{aligned} \tag{36}$$

在计算 (34) 式中的极小值时, 可以将 $\boldsymbol{B}$ 换成 $-\boldsymbol{B}$ 后, 计算极大值的相反数. [来自 Neumann(1937)]. ■

引理 A.4.6 设 $\boldsymbol{P} = (p_{ij})$ 是一个双随机矩阵 $(p_{ij} \geqslant 0, \sum_{i=1}^{p} p_{ij} = 1, \sum_{j=1}^{p} p_{ij} = 1)$. 设 $y_1 \geqslant y_2 \geqslant \cdots \geqslant y_p$, 则

$$\sum_{i=1}^{k} y_i \geqslant \sum_{i=1}^{k}\sum_{j=1}^{n} p_{ij}y_j, \quad k = 1,\cdots,p. \tag{37}$$

证明

$$\sum_{i=1}^{k}\sum_{j=1}^{p} p_{ij}y_j = \sum_{j=1}^{p} g_j y_j, \tag{38}$$

其中 $g_j = \sum_{i=1}^{k} p_{ij}$, $j = 1,\cdots,p$ $(0 \leqslant g_j \leqslant 1, \sum_{j=1}^{p} g_j = k)$. 那么

$$\begin{aligned}\sum_{j=1}^{p} g_j y_j - \sum_{j=1}^{k} y_j &= -\sum_{j=1}^{k} y_j + y_k\left(k - \sum_{j=1}^{p} g_j\right) + \sum_{j=1}^{p} g_j y_j \\ &= \sum_{j=1}^{k}(y_j - y_k)(g_j - 1) + \sum_{j=k+1}^{p}(y_j - y_k)g_j \\ &\leqslant 0.\end{aligned} \tag{39}$$

■

推论 A.4.2 设$\boldsymbol{A}$是对称矩阵, 其特征根为 $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_p$. 那么

$$\max_{R'R=I_k} \text{tr } R'AR = \sum_{i=1}^{k} a_i. \tag{40}$$

证明 在定理 A.4.7 中, 令

$$B = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \quad ■ \tag{41}$$

定理 A.4.8

$$|I + xC| = 1 + x \text{ tr } C + O(x^2). \tag{42}$$

证明 行列式 (42) 是一个关于 x 的 p 次多项式; 一次项的系数是行列式的一阶导数在 $x = 0$ 时的值. 在引理 A.4.5 中令 $n = 1$, $c_1 = x$, $\beta_{ih}(x) = \delta_{ih} + xc_{ih}$, 其中 $\delta_{ii} = 1$ 且 $\delta_{ih} = 0$, $i \neq h$. 从而 $\text{d } \beta_{ih}(x)/\text{d}x = c_{ih}$, 并且当 $x = 0$ 时,$B_{ii} = 1$, 而 $B_{ih} = 0$, $i \neq h$. 因此

$$\left.\frac{\text{d}|B(x)|}{\text{d}x}\right|_{x=0} = \sum_{i=1}^{p} c_{ii}. \quad ■ \tag{43}$$

A.5 Gram-Schmidt 正交化和线性方程组的解

A.5.1 Gram-Schmidt 正交化

在 7.2 节 Wishart 密度的推导过程中, 包含了一组向量的 Gram-Schmidt 正交化, 现在我们回顾一下它的推导过程. 考虑 p 个线性无关的 n 无向量 $v_1 \cdots, v_p$ $(p \leqslant n)$. 定义 $w_1 = v_1$,

$$w_i = v_i - \sum_{j=1}^{i-1} \frac{v_i' w_j}{\| w_j \|^2} w_j, \quad i = 2, \cdots, p. \tag{1}$$

那么 $w_i \neq 0, i = 1, \cdots, p$, 因为 $v_1, \cdots, v_p$ 是线性无关的, 而且正如在 7.2 节的介绍中所证明的那样, $w_i' w_j = 0$, $i \neq j$. 令 $u_i = (1/\| w_i \|) w_i$, $i = 1, \cdots, p$. 那么 $u_1, \cdots, u_p$ 是标准正交的, 也就是说, 它们是相互正交的单位向量. 令 $U = (u_1, \cdots, u_p)$. 则 $U'U = I$. 定义 $t_{ii} = \| w_i \|$ (> 0),

$$t_{ij} = \frac{v_i' w_j}{\| w_j \|} = v_i' u_j, \quad j = 1, \cdots, i-1, \quad i = 2, \cdots, p, \tag{2}$$

且 $t_{ij} = 0$, $j = i+1, \cdots, p$, $i = 1, \cdots, p-1$. 这样 $T = (t_{ij})$ 是一个下三角矩阵. 我们将 (1) 式写成

$$v_i = \| w_j \| u_i + \sum_{j=1}^{i-1} (v_i' u_j) u_j = \sum_{j=1}^{i} t_{ij} u_j, \quad i = 1, \cdots, p, \tag{3}$$

即

$$V = (v_1, \cdots, v_p) = UT'. \tag{4}$$

这样, 如在 7.2 节中所介绍的, 有

$$A = V'V = TU'UT' = TT'. \tag{5}$$

注意到, 如果 V 是方阵, 我们已经可以将任意一个非奇异矩阵分解成为一个正交矩阵和一个上三角矩阵的乘积, 并且这个上三角矩阵的主对角线上的元素都为正数; 这个结论有的时候称作 QR分解. (4) 式中的矩阵 U 和 T 都是唯一的.

这些运算可以按照不同的顺序进行. 令 $V = (v_1^{(0)}, \cdots, v_p^{(0)})$. 对于 $k = 1, \cdots, p-1$, 递归地定义

$$t_{kk}=\| \boldsymbol{v}_k^{(k-1)} \|, \quad \boldsymbol{u}_k=\frac{1}{\| \boldsymbol{v}_k^{(k-1)} \|}\boldsymbol{v}_k^{(k-1)}=\frac{1}{t_{kk}}\boldsymbol{v}_k^{(k-1)}, \tag{6}$$

$$t_{jk}=\boldsymbol{v}_j^{(k-1)\prime}\boldsymbol{u}_k, \quad j=k+1,\cdots,p, \tag{7}$$

$$\boldsymbol{v}_j^{(k)}=\boldsymbol{v}_j^{(k-1)}-t_{jk}\boldsymbol{u}_k, \quad j=k+1,\cdots,p. \tag{8}$$

最后定义 $t_{pp}=\| \boldsymbol{v}_p^{(p-1)} \|$ 以及 $\boldsymbol{u}_p=(1/t_{pp})\boldsymbol{v}_p^{(p-1)}$. 这两个过程给出了同样的正交向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_p$, 以及同样的三角形矩阵 (t_{ij}).

矩阵 $\boldsymbol{V}$ 的列数是任意的. 对于固定的数, 通常最好在每一步中都选择最大的 $\| \boldsymbol{v}_j^{(k-1)} \|$ 来计算 t_{kk}.

我们构造 $\boldsymbol{w}_i$ 时, 并不是使得它与 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{i-1}$ 正交, 而是等价地使得它与 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i-1}$ 正交. 令 $\boldsymbol{w}_1=\boldsymbol{v}_1$, 并且定义

$$\boldsymbol{w}_i=\boldsymbol{v}_i+\sum_{j=1}^{i-1} f_{ij}\boldsymbol{v}_j \tag{9}$$

使得

$$0=\boldsymbol{v}_h'\boldsymbol{w}_i=\boldsymbol{v}_h'\boldsymbol{v}_i+\sum_{j=1}^{i-1} f_{ij}\boldsymbol{v}_h'\boldsymbol{v}_j \tag{10}$$

$$=a_{hi}+\sum_{j=1}^{i-1} a_{hi}f_{ij}, \quad h=1,\cdots,i-1,$$

令 $\boldsymbol{F}=(f_{ij})$, 其中 $f_{ii}=1$ 且 $f_{ij}=0$, $i<j$. 那么

$$\boldsymbol{W}=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_p)=\boldsymbol{V}\boldsymbol{F}'. \tag{11}$$

设 $\boldsymbol{D}_t$ 是对角矩阵, 它的第 j 个对角元为 $\| \boldsymbol{w}_j \|=t_{jj}$. 那么 $\boldsymbol{U}=\boldsymbol{W}\boldsymbol{D}_t^{-1}=\boldsymbol{V}\boldsymbol{F}'\boldsymbol{D}_t^{-1}$. 与 $\boldsymbol{V}=\boldsymbol{U}\boldsymbol{T}'$ 比较, 可知 $\boldsymbol{F}=\boldsymbol{D}\boldsymbol{T}^{-1}$. 因为 $\boldsymbol{A}=\boldsymbol{T}\boldsymbol{T}'$, 我们可以看到 $\boldsymbol{F}\boldsymbol{A}=\boldsymbol{D}\boldsymbol{T}'$ 是上三角矩阵. 因此$\boldsymbol{F}$ 就是定理 A.1.2 中所定义的矩阵.

还有其他的一些方法可以完成 QR 分解, 它们可能是计算上更有效, 或者更稳定的. 一个 Householder矩阵具有形式 $\boldsymbol{H}=\boldsymbol{I}_n-2\boldsymbol{\alpha}\boldsymbol{\alpha}'$, 其中 $\boldsymbol{\alpha}'\boldsymbol{\alpha}=1$, 并且这个矩阵是一个正交的对称矩阵. 可以选择这样一个矩阵 $\boldsymbol{H}_1$ (也就是, 一个向量 $\boldsymbol{\alpha}$), 使得 $\boldsymbol{H}_1\boldsymbol{V}$ 的第一列除了第一个元素为正数外, 其他元素都为 0. 紧接着的矩阵形如

$$\boldsymbol{H}_2=\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-1} \end{pmatrix}-2\begin{pmatrix} 0 \\ \boldsymbol{\alpha} \end{pmatrix}\begin{pmatrix} 0 & \boldsymbol{\alpha}' \end{pmatrix}=\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-1}-2\boldsymbol{\alpha}\boldsymbol{\alpha}' \end{pmatrix}. \tag{12}$$

选择 $(n-1)$ 元向量 $\boldsymbol{\alpha}$, 使得 $\boldsymbol{H}_1\boldsymbol{V}$ 的第二列除了前两个元素为正数外, 其他元素都为 0. 这过程可以一直继续下去, 直到

$$\boldsymbol{H}_{p-1}\cdots\boldsymbol{H}_2\boldsymbol{H}_1\boldsymbol{V}=\begin{bmatrix} \boldsymbol{T}' \\ \mathbf{0} \end{bmatrix}, \tag{13}$$

其中 $\boldsymbol{T}'$ 是上三角矩阵, $\mathbf{0}$ 是 $(n-p)\times p$ 矩阵. 令

$$\boldsymbol{H}'=\boldsymbol{H}_1\cdots\boldsymbol{H}_{p-1}=\begin{pmatrix} \boldsymbol{H}^{(1)} & \boldsymbol{H}^{(2)} \end{pmatrix}, \tag{14}$$

其中 $\boldsymbol{H}^{(1)}$ 有 p 列. 这样由 (13) 我们得到 $\boldsymbol{V}=\boldsymbol{H}^{(1)}\boldsymbol{T}'$. 因为分解是唯一的, $\boldsymbol{H}^{(1)}=\boldsymbol{U}$.

另一种处理程序是吉文斯矩阵. 一个吉文斯矩阵 $\boldsymbol{G}_{ij}$ 除了元素 $g_{ii}=\cos\theta=g_{jj}$ 和 $g_{ij}=\sin\theta=-g_{ji}$, $i\neq j$ 外, 就是单位矩阵 $\boldsymbol{I}$. 它是正交矩阵. 用这个矩阵左乘 $\boldsymbol{V}$, 除了第 i 行和第 j 行之外, 其他所有行都没有变化. 可以选择 θ, 使得 $\boldsymbol{G}_{ij}\boldsymbol{V}$ 的第 (i,j) 元为 0. 依次选

择吉文斯矩阵 $\boldsymbol{G}_{21},\cdots,\boldsymbol{G}_{n1}$, 使得 $\boldsymbol{G}_{n1}\cdots\boldsymbol{G}_{21}\boldsymbol{V}$ 的第一列中除了第一个元素为正数外其他的都为 0. 接下来选择 $\boldsymbol{G}_{32},\cdots,\boldsymbol{G}_{n2}$, 使得当它们作用之后, 得到的矩阵中除了第二列中前两个元素为正数外, 其他的元素都为 0. 记

$$\boldsymbol{G}'=\boldsymbol{G}'_{21}\cdots\boldsymbol{G}'_{n1}\boldsymbol{G}'_{32}\cdots\boldsymbol{G}'_{n,p-1}=\begin{pmatrix}\boldsymbol{G}^{(1)} & \boldsymbol{G}^{(2)}\end{pmatrix}. \tag{15}$$

那么, 我们得到

$$\boldsymbol{V}=\boldsymbol{G}'\begin{bmatrix}\boldsymbol{T}'\\ \boldsymbol{0}\end{bmatrix}=\boldsymbol{G}^{(1)}\boldsymbol{T}', \tag{16}$$

且 $\boldsymbol{G}^{(1)}=\boldsymbol{U}$.

A.5.2　线性方程组的解

在计算回归系数和其他统计量时, 我们需要求解线性方程组

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}, \tag{17}$$

其中 $\boldsymbol{A}$ 是 $p\times p$ 的正定矩阵. 求解的一个方法是高斯变量消去法, 或称枢轴凝聚法. 在定理 A.1.2 的证明中我们构造了一个对角线上的元素为 1 的下三角矩阵 $\boldsymbol{F}$, 使得 $\boldsymbol{F}\boldsymbol{A}=\boldsymbol{A}^*$ 是一个上三角矩阵. 如果 $\boldsymbol{F}\boldsymbol{y}=\boldsymbol{y}^*$, 那么方程组为

$$\boldsymbol{A}^*\boldsymbol{x}=\boldsymbol{y}^*. \tag{18}$$

用坐标表示就是

$$\sum_{j=1}^{p}a_{ij}^*x_j=y_i^*. \tag{19}$$

令 $a_{ij}^{**}=a_{ij}^*/a_{ii}^*,\ y_i^{**}=y_i^*/a_{ii}^*,\ j=i,i+1,\cdots,p,\ i=1,\cdots,p$. 那么

$$x_i=y_i^{**}-\sum_{j=i+1}^{p}a_{ij}^{**}x_i; \tag{20}$$

从这些方程组可以连续求解出 $x_p,x_{p-1},\cdots,x_1$. 计算 $\boldsymbol{F}\boldsymbol{A}=\boldsymbol{A}^*$ 的方法称为*向前解法*, 而 (18) 式的方法称为*向后解法*.

因为 $\boldsymbol{F}\boldsymbol{A}\boldsymbol{F}'=\boldsymbol{A}^*\boldsymbol{F}'=\boldsymbol{D}^2$ 是对角矩阵, (20) 式为 $\boldsymbol{A}^{**}\boldsymbol{x}=\boldsymbol{y}^{**}$, 其中 $\boldsymbol{A}^{**}=\boldsymbol{D}^{-2}\boldsymbol{A}^*$ 且 $\boldsymbol{y}^{**}=\boldsymbol{D}^{-2}\boldsymbol{y}^*$. 求解这个方程组得到

$$\boldsymbol{x}=\boldsymbol{A}^{**-1}\boldsymbol{y}^{**}=\boldsymbol{F}'\boldsymbol{y}^{**}. \tag{21}$$

计算结果为

$$\boldsymbol{x}=\boldsymbol{F}_1'\cdots\boldsymbol{F}_{p-1}'\boldsymbol{D}^{-2}\boldsymbol{F}_{p-1}\cdots\boldsymbol{F}_1\boldsymbol{y}. \tag{22}$$

在 (22) 中与 $\boldsymbol{y}$ 相乘的部分, 表示的是生成 $\boldsymbol{A}^{-1}$ 所需要的一系列行变换.

向前解法的操作是将 $\boldsymbol{A}$ 变换成一个上三角矩阵 $\boldsymbol{A}^*$. 正如在 A.5.1 节所看到的, 一个矩阵的三角化可以用一系列 Householder 变换或者吉文斯变换来完成.

由 $\boldsymbol{F}\boldsymbol{A}=\boldsymbol{A}^*$, 我们得到

$$|\boldsymbol{A}|=\prod_{i=1}^{p}a_{ii}^{(i)}, \tag{23}$$

由向前解法的结果知, 它正是 $\boldsymbol{A}^*$ 的对角线上的元素的乘积. 我们也得到

$$\begin{aligned}\boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{y}&=(\boldsymbol{F}\boldsymbol{y})'\boldsymbol{D}^{-2}(\boldsymbol{F}\boldsymbol{y})=\boldsymbol{y}^{*\prime}\boldsymbol{D}^{-2}\boldsymbol{y}^*\\&=\boldsymbol{y}^{*\prime}\boldsymbol{y}^{**}.\end{aligned} \tag{24}$$

向前解法给出了一个计算在 T^2 中出现的二次型以及其他统计量的算法.

关于矩阵计算的更多结果, 请参考 Golub and Von Loan(1989).

附 录 B 表

表 B.1 Wilk 似然准则：因子 $C(p,m,M)$ 对 $\chi^2_{p,m}$ 的调整，其中 $M=n-p+1$

	5% 显著性水平								
					$p=3$				
$M\backslash m$	2	4	6	8	10	12	14	16	18
1	1.295	1.422	1.535	1.632	1.716	1.791	1.857	1.916	1.971
2	1.109	1.174	1.241	1.302	1.359	1.410	1.458	1.501	1.542
3	1.058	1.099	1.145	1.190	1.232	1.272	1.309	1.344	1.377
4	1.036	1.065	1.099	1.133	1.167	1.199	1.229	1.258	1.286
5	1.025	1.046	1.072	1.100	1.127	1.154	1.179	1.204	1.228
6	1.018	1.035	1.056	1.078	1.101	1.123	1.145	1.167	1.188
7	1.014	1.027	1.044	1.063	1.082	1.101	1.121	1.139	1.158
8	1.011	1.022	1.036	1.052	1.068	1.085	1.102	1.119	1.135
9	1.009	1.018	1.030	1.043	1.058	1.073	1.088	1.102	1.117
10	1.007	1.015	1.025	1.037	1.050	1.063	1.076	1.089	1.103
12	1.005	1.011	1.019	1.028	1.038	1.048	1.059	1.070	1.081
15	1.003	1.008	1.013	1.020	1.027	1.035	1.043	1.052	1.060
20	1.002	1.004	1.008	1.012	1.017	1.022	1.028	1.034	1.040
30	1.001	1.002	1.004	1.006	1.009	1.011	1.015	1.018	1.021
60	1.000	1.001	1.001	1.002	1.002	1.003	1.004	1.006	1.007
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	12.5916	21.0261	28.8693	36.4150	43.7730	50.9985	58.1240	65.1708	72.1532

	5% 显著性水平								
	$p=3$		$p=4$						
$M\backslash m$	20	22	2	4	6	8	10	12	14
1	2.021	2.067	1.407	1.451	1.517	1.583	1.644	1.700	1.751
2	1.580	1.616	1.161	1.194	1.240	1.286	1.331	1.373	1.413
3	1.408	1.438	1.089	1.114	1.148	1.183	1.218	1.252	1.284
4	1.313	1.338	1.057	1.076	1.102	1.130	1.159	1.186	1.213
5	1.251	1.273	1.040	1.055	1.076	1.099	1.122	1.145	1.168
6	1.208	1.227	1.030	1.042	1.059	1.078	1.097	1.118	1.137
7	1.176	1.193	1.023	1.033	1.047	1.063	1.080	1.097	1.115
8	1.151	1.167	1.018	1.027	1.038	1.052	1.067	1.082	1.097
9	1.132	1.147	1.015	1.022	1.032	1.044	1.057	1.070	1.084
10	1.116	1.129	1.012	1.018	1.027	1.038	1.049	1.061	1.073
12	1.092	1.103	1.009	1.014	1.020	1.029	1.038	1.047	1.058
15	1.069	1.078	1.006	1.009	1.014	1.020	1.027	1.035	1.042
20	1.046	1.052	1.003	1.006	1.009	1.013	1.017	1.022	1.027
30	1.025	1.029	1.002	1.003	1.004	1.006	1.009	1.011	1.014
60	1.008	1.009	1.000	1.001	1.001	1.002	1.003	1.003	1.004
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	79.0819	85.9649	15.5073	26.2962	36.4150	46.1943	55.7585	65.1708	74.4683

(续)

$M\backslash m$	5% 显著性水平 $p=4$ 16	18	20	$p=5$ 2	4	6	8	10	12
1	1.799	1.843	1.884	1.503	1.483	1.514	1.556	1.600	1.643
2	1.450	1.485	1.518	1.209	1.216	1.245	1.280	1.315	1.350
3	1.314	1.343	1.371	1.120	1.130	1.154	1.182	1.211	1.240
4	1.239	1.264	1.288	1.079	1.089	1.108	1.131	1.155	1.179
5	1.190	1.212	1.233	1.056	1.065	1.081	1.100	1.120	1.141
6	1.157	1.176	1.194	1.042	1.050	1.063	1.079	1.097	1.114
7	1.132	1.149	1.165	1.033	1.040	1.051	1.065	1.080	1.095
8	1.113	1.128	1.143	1.026	1.032	1.042	1.054	1.067	1.081
9	1.098	1.111	1.125	1.022	1.027	1.035	1.046	1.057	1.070
10	1.086	1.098	1.110	1.018	1.023	1.030	1.039	1.050	1.061
12	1.068	1.078	1.088	1.013	1.017	1.023	1.030	1.038	1.047
15	1.050	1.058	1.066	1.009	1.011	1.016	1.021	1.028	1.034
20	1.033	1.039	1.045	1.005	1.007	1.010	1.013	1.018	1.022
30	1.018	1.021	1.024	1.002	1.003	1.005	1.007	1.009	1.012
60	1.005	1.007	1.008	1.001	1.001	1.001	1.002	1.003	1.004
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	83.6753	92.8083	101.879	18.3070	31.4104	43.7730	55.7585	67.5048	79.0819

$M\backslash m$	5% 显著性水平 $p=5$ 14	16	$p=6$ 2	6	8	10	12	$p=7$ 2	4
1	1.683	1.722	1.587	1.520	1.543	1.573	1.605	1.662	1.550
2	1.383	1.415	1.254	1.255	1.279	1.307	1.335	1.297	1.263
3	1.267	1.294	1.150	1.163	1.184	1.208	1.232	1.178	1.165
4	1.203	1.226	1.100	1.116	1.134	1.154	1.175	1.121	1.116
5	1.161	1.181	1.072	1.088	1.103	1.120	1.138	1.089	1.087
6	1.132	1.150	1.055	1.069	1.082	1.097	1.113	1.068	1.068
7	1.111	1.127	1.043	1.056	1.068	1.081	1.095	1.054	1.055
8	1.095	1.109	1.035	1.046	1.057	1.068	1.081	1.044	1.045
9	1.082	1.095	1.029	1.039	1.048	1.059	1.070	1.036	1.038
10	1.072	1.083	1.024	1.034	1.042	1.051	1.061	1.031	1.032
12	1.057	1.066	1.018	1.025	1.032	1.040	1.048	1.023	1.024
15	1.042	1.049	1.012	1.018	1.023	1.029	1.035	1.016	1.017
20	1.027	1.033	1.007	1.011	1.014	1.018	1.023	1.010	1.011
30	1.014	1.018	1.003	1.006	1.007	1.010	1.012	1.005	1.005
60	1.004	1.006	1.001	1.002	1.002	1.003	1.004	1.001	1.001
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	90.5312	101.879	21.0261	50.9985	65.1708	79.0819	92.8083	23.6848	41.3371

(续)

$M\backslash m$	5% 显著性水平								
	$p=7$			$p=8$		$p=9$			$p=10$
	6	8	10	2	8	2	4	6	2
1	1.530	1.538	1.557	1.729	1.538	1.791	1.614	1.558	1.847
2	1.266	1.282	1.303	1.336	1.288	1.373	1.309	1.293	1.408
3	1.173	1.189	1.208	1.206	1.195	1.232	1.201	1.196	1.257
4	1.124	1.139	1.155	1.142	1.144	1.162	1.144	1.144	1.182
5	1.095	1.108	1.122	1.105	1.113	1.121	1.110	1.112	1.137
6	1.075	1.086	1.099	1.081	1.091	1.094	1.088	1.090	1.107
7	1.062	1.071	1.083	1.065	1.076	1.076	1.071	1.074	1.087
8	1.051	1.060	1.070	1.053	1.064	1.062	1.060	1.062	1.072
9	1.043	1.051	1.060	1.044	1.055	1.052	1.050	1.053	1.061
10	1.037	1.044	1.053	1.038	1.048	1.045	1.043	1.046	1.052
12	1.029	1.034	1.042	1.028	1.038	1.034	1.033	1.035	1.039
15	1.020	1.024	1.031	1.019	1.027	1.023	1.023	1.025	1.028
20	1.013	1.016	1.019	1.012	1.017	1.014	1.015	1.016	1.017
30	1.006	1.008	1.010	1.006	1.009	1.007	1.007	1.008	1.009
60	1.002	1.002	1.003	1.001	1.003	1.002	1.002	1.002	1.002
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	58.1240	74.4683	90.5312	26.2962	83.6753	28.8693	50.9985	72.1532	31.4104

$M\backslash m$	1% 显著性水平							
	$p=3$							
	2	4	6	8	10	12	14	16
1	1.356	1.514	1.649	1.763	1.862	1.949	2.026	2.095
2	1.131	1.207	1.282	1.350	1.413	1.470	1.523	1.571
3	1.070	1.116	1.167	1.216	1.262	1.306	1.346	1.384
4	1.043	1.076	1.113	1.150	1.187	1.221	1.254	1.285
5	1.030	1.054	1.082	1.112	1.141	1.170	1.198	1.224
6	1.022	1.040	1.063	1.087	1.112	1.136	1.159	1.182
7	1.016	1.031	1.050	1.070	1.091	1.111	1.132	1.152
8	1.013	1.025	1.041	1.058	1.075	1.093	1.111	1.129
9	1.010	1.021	1.034	1.048	1.064	1.080	1.095	1.111
10	1.009	1.017	1.028	1.041	1.055	1.069	1.082	1.097
12	1.006	1.012	1.021	1.031	1.042	1.053	1.064	1.076
15	1.004	1.009	1.014	1.021	1.030	1.038	1.047	1.056
20	1.002	1.005	1.009	1.013	1.019	1.024	1.030	1.036
30	1.001	1.002	1.004	1.007	1.009	1.012	1.016	1.019
60	1.000	1.001	1.001	1.002	1.003	1.004	1.005	1.006
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	16.8119	26.2170	34.8053	42.9798	50.8922	58.6192	66.2062	73.6826

(续)

$M\backslash m$	1% 显著性水平							
	$p=3$			$p=4$				
	18	20	22	2	4	6	8	10
1	2.158	2.216	2.269	1.490	1.550	1.628	1.704	1.774
2	1.616	1.657	1.696	1.192	1.229	1.279	1.330	1.379
3	1.420	1.453	1.485	1.106	1.132	1.168	1.207	1.244
4	1.315	1.344	1.371	1.068	1.088	1.115	1.146	1.176
5	1.249	1.274	1.297	1.047	1.063	1.085	1.109	1.134
6	1.204	1.226	1.246	1.035	1.048	1.066	1.086	1.107
7	1.171	1.190	1.209	1.027	1.037	1.052	1.070	1.088
8	1.146	1.163	1.180	1.021	1.030	1.043	1.058	1.073
9	1.127	1.142	1.157	1.017	1.025	1.036	1.048	1.062
10	1.111	1.125	1.139	1.014	1.021	1.030	1.041	1.054
12	1.087	1.099	1.110	1.010	1.015	1.023	1.031	1.041
15	1.065	1.074	1.083	1.007	1.010	1.016	1.022	1.029
20	1.043	1.049	1.056	1.004	1.006	1.010	1.014	1.019
30	1.023	1.027	1.031	1.002	1.003	1.005	1.007	1.009
60	1.007	1.009	1.010	1.000	1.001	1.001	1.002	1.003
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	81.0688	88.3794	95.6257	20.0902	31.9999	42.9798	53.4858	63.6907

$M\backslash m$	1% 显著性水平							
	$p=4$					$p=5$		
	12	14	16	18	20	2	4	6
1	1.838	1.896	1.949	1.999	2.045	1.606	1.589	1.625
2	1.424	1.467	1.507	1.545	1.580	1.248	1.253	1.284
3	1.280	1.314	1.347	1.378	1.408	1.141	1.150	1.175
4	1.205	1.234	1.261	1.287	1.313	1.092	1.101	1.121
5	1.159	1.183	1.207	1.230	1.252	1.065	1.074	1.090
6	1.128	1.149	1.169	1.189	1.208	1.049	1.056	1.070
7	1.106	1.124	1.142	1.160	1.177	1.038	1.044	1.056
8	1.089	1.105	1.121	1.137	1.153	1.031	1.036	1.046
9	1.076	1.091	1.105	1.119	1.133	1.025	1.030	1.039
10	1.066	1.079	1.092	1.105	1.118	1.021	1.025	1.033
12	1.051	1.062	1.073	1.083	1.094	1.015	1.019	1.025
15	1.037	1.045	1.053	1.062	1.071	1.010	1.013	1.017
20	1.024	1.029	1.035	1.041	1.047	1.006	1.008	1.011
30	1.012	1.015	1.019	1.022	1.026	1.003	1.004	1.005
60	1.004	1.005	1.006	1.007	1.008	1.001	1.001	1.001
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	73.6826	83.5134	93.2168	102.8168	112.3292	23.2093	37.5662	50.8922

(续)

M\m	1% 显著性水平							
	p = 5					p = 6		
	8	10	12	14	16	2	6	8
1	1.672	1.721	1.768	1.813	1.855	1.707	1.631	1.656
2	1.321	1.359	1.396	1.431	1.465	1.300	1.294	1.319
3	1.204	1.235	1.265	1.294	1.323	1.175	1.183	1.205
4	1.145	1.171	1.196	1.221	1.245	1.116	1.129	1.148
5	1.110	1.131	1.153	1.174	1.196	1.084	1.097	1.113
6	1.087	1.105	1.124	1.143	1.161	1.063	1.076	1.090
7	1.071	1.087	1.103	1.119	1.136	1.050	1.061	1.074
8	1.059	1.073	1.087	1.102	1.116	1.040	1.051	1.062
9	1.050	1.062	1.075	1.088	1.101	1.033	1.043	1.052
10	1.043	1.054	1.065	1.077	1.089	1.028	1.037	1.045
12	1.033	1.041	1.051	1.060	1.070	1.021	1.028	1.035
15	1.023	1.030	1.037	1.044	1.052	1.014	1.020	1.024
20	1.015	1.019	1.024	1.029	1.034	1.008	1.012	1.015
30	1.007	1.010	1.012	1.015	1.019	1.004	1.006	1.008
60	1.002	1.003	1.004	1.005	1.006	1.001	1.002	1.002
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	63.6907	76.1539	88.3794	100.4252	112.3299	26.2170	58.6192	73.6826

M\m	1% 显著性水平						
	p = 6		p = 7				
	10	12	2	4	6	8	10
1	1.687	1.722	1.797	1.667	1.642	1.648	1.666
2	1.348	1.378	1.348	1.305	1.306	1.321	1.342
3	1.230	1.255	1.207	1.188	1.194	1.210	1.229
4	1.169	1.191	1.140	1.130	1.138	1.152	1.169
5	1.131	1.150	1.102	1.097	1.105	1.117	1.132
6	1.106	1.122	1.078	1.076	1.083	1.094	1.107
7	1.087	1.102	1.062	1.061	1.067	1.077	1.089
8	1.074	1.086	1.050	1.050	1.056	1.065	1.075
9	1.063	1.075	1.042	1.042	1.047	1.055	1.065
10	1.055	1.065	1.035	1.036	1.041	1.048	1.056
12	1.042	1.051	1.026	1.027	1.031	1.037	1.044
15	1.030	1.037	1.018	1.019	1.022	1.025	1.032
20	1.019	1.024	1.011	1.012	1.014	1.017	1.020
30	1.010	1.013	1.005	1.006	1.007	1.009	1.011
60	1.003	1.004	1.001	1.002	1.002	1.003	1.003
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	88.3794	102.816	29.1412	48.2782	66.2062	83.5134	100.425

(续)

	1% 显著性水平					
	$p=8$		$p=9$			$p=10$
$M\backslash m$	2	8	2	4	6	2
1	1.879	1.646	1.953	1.740	1.671	2.021
2	1.394	1.326	1.436	1.355	1.333	1.476
3	1.238	1.215	1.267	1.226	1.218	1.296
4	1.163	1.158	1.185	1.161	1.158	1.207
5	1.120	1.123	1.138	1.122	1.122	1.155
6	1.092	1.099	1.107	1.096	1.098	1.121
7	1.074	1.082	1.086	1.078	1.080	1.098
8	1.060	1.069	1.070	1.065	1.067	1.081
9	1.050	1.059	1.059	1.055	1.058	1.068
10	1.043	1.051	1.050	1.047	1.050	1.058
12	1.032	1.040	1.038	1.036	1.037	1.044
15	1.022	1.028	1.026	1.026	1.027	1.031
20	1.013	1.018	1.016	1.016	1.017	1.018
30	1.007	1.009	1.008	1.008	1.009	1.010
60	1.002	1.003	1.002	1.002	1.003	1.003
∞	1.000	1.000	1.000	1.000	1.000	1.000
χ^2_{pm}	31.9999	93.2168	34.8053	58.6192	81.0688	37.5662

表 B.2 Lawley-Hotelling 迹检验的分位数表

$$\Pr\left\{\frac{n}{m}W \geqslant x_\alpha\right\} = \alpha$$

	5% 显著性水平								
	$p=2$								
$n\backslash m$	2	3	4	5	6	8	10	12	15
2	9.859*	10.659*	11.098*	11.373*	11.562*	11.804*	11.952*	12.052*	12.153*
3	58.428	58.915	59.161	59.308	59.407	59.531	59.606	59.655	59.705
4	23.999	23.312	22.918	22.663	22.484	22.250	22.104	22.003	21.901
5	15.639	14.864	14.422	14.135	13.934	13.670	13.504	13.391	13.275
6	12.175	11.411	10.975	10.691	10.491	10.228	10.063	9.949	9.832
7	10.334	9.594	9.169	8.893	8.697	8.440	8.277	8.164	8.048
8	9.207	8.488	8.075	7.805	7.614	7.361	7.201	7.090	6.975
10	7.909	7.224	6.829	6.570	6.386	6.141	5.984	5.875	5.761
12	7.190	6.528	6.146	5.894	5.715	5.474	5.320	5.212	5.100
14	6.735	6.090	5.717	5.470	5.294	5.057	4.905	4.798	4.686
18	6.193	5.571	5.209	4.970	4.798	4.566	4.416	4.309	4.198
20	6.019	5.405	5.047	4.810	4.640	4.410	4.260	4.154	4.042
25	5.724	5.124	4.774	4.542	4.374	4.147	3.998	3.892	3.780
30	5.540	4.949	4.604	4.374	4.209	3.983	3.835	3.729	3.617
35	5.414	4.829	4.488	4.260	4.096	3.872	3.724	3.618	3.505
40	5.322	4.742	4.404	4.178	4.014	3.791	3.643	3.538	3.425
50	5.198	4.625	4.290	4.066	3.904	3.682	3.535	3.429	3.315
60	5.118	4.549	4.217	3.994	3.833	3.611	3.465	3.359	3.245
70	5.062	4.496	4.165	3.944	3.783	3.562	3.416	3.310	3.196
80	5.020	4.457	4.127	3.907	3.747	3.526	3.380	3.274	3.159
100	4.963	4.403	4.075	3.856	3.696	3.476	3.330	3.224	3.109
200	4.851	4.298	3.974	3.757	3.598	3.380	3.234	3.127	3.012
∞	4.744	4.197	3.877	3.661	3.504	3.287	3.141	3.035	2.918

* 乘 10^2.

(续)

	1% 显著性水平								
	$p=2$								
$n\backslash m$	2	3	4	5	6	8	10	12	15
2	2.467†	2.667†	2.776†	2.844†	2.891†	2.952†	2.989†	3.014†	3.039†
3	2.985*	2.990*	2.992*	2.994*	2.995*	2.996*	2.997*	2.997*	2.998*
4	74.275	71.026	69.244	68.116	67.337	66.332	65.712	65.290	64.862
5	38.295	35.567	34.070	33.121	32.465	31.615	31.088	30.729	30.364
6	26.118	23.794	22.517	21.706	21.143	20.413	19.958	19.648	19.332
7	20.388	18.326	17.191	16.469	15.967	15.313	14.905	14.626	14.341
8	17.152	15.268	14.229	13.567	13.106	12.504	12.127	11.868	11.603
10	13.701	12.038	11.120	10.531	10.121	9.582	9.243	9.010	8.769
12	11.920	10.388	9.541	8.996	8.615	8.113	7.796	7.577	7.351
14	10.844	9.399	8.597	8.082	7.720	7.242	6.939	6.729	6.511
18	9.617	8.278	7.533	7.053	6.714	6.265	5.979	5.780	5.572
20	9.236	7.932	7.206	6.736	6.406	5.966	5.685	5.489	5.284
25	8.604	7.360	6.666	6.217	5.899	5.476	5.204	5.013	4.813
30	8.219	7.013	6.339	5.903	5.593	5.180	4.914	4.726	4.529
35	7.959	6.780	6.120	5.692	5.389	4.982	4.720	4.535	4.339
40	7.773	6.613	5.964	5.542	5.243	4.841	4.582	4.398	4.204
50	7.523	6.389	5.754	5.341	5.048	4.653	4.397	4.216	4.023
60	7.363	3.247	5.621	5.214	4.924	4.534	4.280	4.100	3.908
70	7.252	6.148	5.529	5.125	4.838	4.451	4.199	4.020	3.829
80	7.171	6.075	5.461	5.061	4.775	4.391	4.140	3.961	3.770
100	7.059	5.976	5.369	4.972	4.690	4.308	4.059	3.881	3.691
200	6.843	5.785	5.191	4.803	4.525	4.150	3.903	3.727	3.538
∞	6.638	5.604	5.023	4.642	4.369	4.000	3.757	3.582	3.393

† 乘 10^4.

* 乘 10^2.

(续)

$n\backslash m$	3	4	5	6	8	10	12	15	20
	5% 显著性水平								
	$p=3$								
3	25.930*	26.996*	27.665*	28.125*	28.712*	29.073*	29.316*	29.561*	29.809*
4	1.188*	1.193*	1.196*	1.198*	1.200*	1.202*	1.203*	1.204*	1.205*
5	42.474	41.764	41.305	40.983	40.562	40.300	40.120	39.937	39.750
6	25.456	24.715	24.235	23.899	23.458	23.182	22.992	22.799	22.600
7	18.752	18.056	17.605	17.288	16.870	16.608	16.427	16.241	16.051
8	15.308	14.657	14.233	13.934	13.540	13.290	13.118	12.941	12.758
10	11.893	11.306	10.921	10.649	10.287	10.057	9.897	9.732	9.560
12	10.229	9.682	9.323	9.068	8.727	8.509	8.357	8.198	8.033
14	9.255	8.736	8.394	8.149	7.822	7.612	7.465	7.311	7.150
16	8.618	8.118	7.788	7.553	7.236	7.031	6.887	6.736	6.577
18	8.170	7.685	7.364	7.135	6.825	6.624	6.483	6.334	6.177
20	7.838	7.365	7.051	6.826	6.522	6.325	6.185	6.038	5.882
25	7.294	6.841	6.539	6.323	6.029	5.837	5.700	5.556	5.401
30	6.965	6.524	6.231	6.020	5.732	5.543	5.409	5.265	5.112
35	6.745	6.313	6.025	5.818	5.534	5.348	5.214	5.072	4.919
40	6.588	6.162	5.878	5.673	5.393	5.208	5.076	4.934	4.781
50	6.377	5.961	5.682	5.481	5.205	5.022	4.891	4.750	4.597
60	6.243	5.832	5.558	5.359	5.086	4.904	4.774	4.633	4.480
70	6.150	5.744	5.471	5.274	5.003	4.823	4.693	4.553	4.399
80	6.082	5.679	5.408	5.212	4.943	4.763	4.634	4.493	4.339
100	5.989	5.590	5.322	5.128	4.860	4.682	4.552	4.413	4.258
200	5.810	5.419	5.156	4.965	4.702	4.525	4.397	4.257	4.102
∞	5.640	5.256	4.999	4.812	4.552	4.377	4.250	4.110	3.954

* 乘 10^2.

(续)

	1% 显著性水平								
	$p=3$								
$n\backslash m$	3	4	5	6	8	10	12	15	20
3	6.484†	6.750†	6.917†	7.031†	7.178†	7.267†	7.328†	7.389†	7.451†
4	5.990*	5.995*	5.998*	6.000*	6.002*	6.003*	6.005*	6.006*	6.007*
5	1.274*	1.242*	1.222*	1.208*	1.190*	1.179*	1.172*	1.164*	1.156*
6	59.507	57.032	55.462	54.377	52.973	52.102	51.509	50.906	50.292
7	37.994	35.993	34.721	33.840	32.695	31.984	31.498	31.002	30.496
8	28.308	26.599	25.511	24.755	23.771	23.157	22.737	22.308	21.868
10	19.737	18.355	17.471	16.855	16.050	15.544	15.197	14.840	14.472
12	15.973	14.765	13.990	13.448	12.737	12.288	11.978	11.659	11.328
14	13.905	12.803	12.096	11.599	10.945	10.530	10.243	9.946	9.638
16	12.610	11.581	10.918	10.452	9.836	9.444	9.172	8.890	8.596
18	11.729	10.751	10.120	9.676	9.087	8.712	8.450	8.178	7.893
20	11.091	10.152	9.545	9.117	8.549	8.186	7.932	7.668	7.390
25	10.075	9.201	8.634	8.233	7.699	7.356	7.115	6.803	6.596
30	9.479	8.644	8.102	7.718	7.205	6.874	6.641	6.395	6.135
35	9.087	8.280	7.755	7.382	6.883	6.560	6.332	6.091	5.834
40	8.811	8.023	7.511	7.146	6.656	6.339	6.115	5.877	5.623
50	8.448	7.686	7.189	6.836	6.360	6.050	5.831	5.597	5.346
60	8.220	7.474	6.988	6.642	6.174	5.870	5.653	5.422	5.172
70	8.063	7.329	6.850	6.509	6.047	5.746	5.531	5.302	5.053
80	7.948	7.224	6.750	6.412	5.955	5.656	5.443	5.215	4.967
100	7.793	7.081	6.614	6.281	5.830	5.534	5.323	5.096	4.850
200	7.498	6.808	6.356	6.032	5.593	5.304	5.096	4.873	4.627
∞	7.222	6.554	6.116	5.801	5.373	5.089	4.885	4.664	4.419

† 乘 10^4.

* 乘 10^2.

(续)

$n\backslash m$	5% 显著性水平								
	$p=4$								
	4	5	6	8	10	12	15	20	25
4	49.964*	51.204*	52.054*	53.142*	53.808*	54.258*	54.71*	55.17*	55.46*
5	1.996*	2.001*	2.005*	2.009*	2.011*	2.013*	2.015*	2.016*	2.017*
6	65.715	64.999	64.497	63.841	63.432	63.151	62.866	62.573	62.396
7	37.343	36.629	36.129	35.474	35.064	34.782	34.495	34.200	34.019
8	26.516	25.868	25.413	24.814	24.437	24.178	23.912	23.639	23.471
10	17.875	17.326	16.938	16.424	16.098	15.872	15.640	15.399	15.250
12	14.338	13.848	13.500	13.037	12.741	12.535	12.321	12.099	11.961
14	12.455	12.002	11.680	11.248	10.972	10.778	10.577	10.366	10.234
16	11.295	10.868	10.563	10.154	9.890	9.705	9.512	9.309	9.181
18	10.512	10.104	9.812	9.419	9.165	8.986	8.798	8.600	8.475
20	9.950	9.556	9.274	8.893	8.645	8.471	8.287	8.093	7.970
25	9.059	8.688	8.422	8.062	7.826	7.659	7.482	7.293	7.173
30	8.538	8.182	7.927	7.578	7.350	7.188	7.015	6.829	6.710
35	8.197	7.852	7.603	7.263	7.040	6.880	6.710	6.526	6.408
40	7.957	7.619	7.375	7.041	6.821	6.664	6.495	6.313	6.195
50	7.640	7.313	7.075	6.750	6.535	6.380	6.214	6.033	5.916
60	7.442	7.120	6.887	6.568	6.356	6.203	6.038	5.858	5.740
70	7.305	6.988	6.758	6.443	6.232	6.081	5.917	5.738	5.620
80	7.206	6.892	6.665	6.351	6.143	5.992	5.829	5.650	5.532
100	7.071	6.762	6.537	6.228	6.021	5.872	5.710	5.531	5.413
200	6.814	6.514	6.295	5.993	5.791	5.644	5.484	5.305	5.186
∞	6.574	6.282	6.069	5.774	5.576	5.431	5.272	5.094	4.974

* 乘 10^2.

(续)

	1% 显著性水平								
	$p=4$								
$n\backslash m$	4	5	6	8	10	12	15	20	25
4	12.491†	12.800†	13.012†	13.283†	13.449†	13.561†	13.67†	13.79†	13.87†
5	9.999*	10.004*	10.008*	10.012*	10.014*	10.016*	10.018*	10.02*	10.02*
6	1.938*	1.906*	1.885*	1.857*	1.840*	1.828*	1.816*	1.804*	1.797*
7	85.053	82.731	81.125	79.047	77.759	76.882	75.989	75.082	74.522
8	51.991	50.178	48.921	47.290	46.276	45.583	44.877	44.156	43.715
10	29.789	28.478	27.566	26.376	25.632	25.121	24.597	24.060	23.731
12	21.965	20.889	20.138	19.154	18.534	18.108	17.668	17.215	16.936
14	18.142	17.199	16.539	15.670	15.121	14.742	14.349	13.943	13.691
16	15.916	15.059	14.457	13.662	13.157	12.807	12.444	12.066	11.831
18	14.473	13.674	13.112	12.368	11.894	11.564	11.221	10.863	10.639
20	13.466	12.710	12.177	11.470	11.018	10.703	10.374	10.030	9.814
25	11.924	11.237	10.751	10.103	9.687	9.395	9.089	8.766	8.562
30	11.055	10.409	9.951	9.338	8.943	8.665	8.372	8.060	7.863
35	10.499	9.880	9.440	8.851	8.470	8.200	7.915	7.611	7.418
40	10.114	9.514	9.087	8.514	8.142	7.879	7.600	7.301	7.110
50	9.614	9.040	8.631	8.079	7.720	7.465	7.194	6.902	6.713
60	9.305	8.747	8.319	7.811	7.460	7.210	6.943	6.655	6.468
70	9.095	8.549	8.158	7.630	7.284	7.037	6.774	6.488	6.301
80	8.944	8.405	8.020	7.498	7.157	6.912	6.651	6.367	6.181
100	8.739	8.211	7.833	7.321	6.985	6.744	6.486	6.204	6.019
200	8.354	7.848	7.484	6.990	6.664	6.429	6.176	5.898	5.714
∞	8.000	7.513	7.163	6.686	6.369	6.140	5.892	5.616	5.432

† 乘 10^4.

* 乘 10^2.

(续)

n\m	5% 显著性水平								
	p = 5								
	5	6	8	10	12	15	20	25	40
5	81.991*	83.352	85.093	86.160†	86.88†	–	–	–	–
6	3.009*	3.014*	3.020*	3.024†	3.027†	3.029†	3.032†	–	–
7	93.762	93.042	92.102	91.515	91.113	90.705	90.29	90.04	–
8	51.339	50.646	49.739	49.170	48.780	48.382	47.973	47.723	47.35
10	27.667	27.115	26.387	25.927	25.610	25.284	24.947	24.740	24.422
12	20.169	19.701	19.079	18.683	18.409	18.124	17.830	17.647	17.365
14	16.643	16.224	15.666	15.309	15.059	14.800	14.530	14.361	14.100
16	14.624	14.239	13.722	13.389	13.157	12.914	12.659	12.499	12.250
18	13.326	12.963	12.476	12.161	11.939	11.708	11.463	11.310	11.068
20	12.424	12.078	11.612	11.310	11.097	10.874	10.637	10.488	10.252
25	11.046	10.728	10.297	10.016	9.817	9.606	9.381	9.239	9.010
30	10.270	9.969	9.559	9.291	9.099	8.896	8.679	8.539	8.314
35	9.774	9.484	9.088	8.828	8.642	8.444	8.230	8.093	7.869
40	9.429	9.147	8.761	8.507	8.325	8.130	7.919	7.783	7.561
50	8.982	8.711	8.339	8.092	7.915	7.725	7.518	7.383	7.161
60	8.706	7.441	8.077	7.836	7.662	7.474	7.269	7.135	6.912
70	8.517	8.257	7.899	7.661	7.489	7.304	7.100	6.967	6.743
80	8.381	8.124	7.770	7.535	7.365	7.181	6.978	6.845	6.621
100	8.197	7.945	7.597	7.365	7.197	7.014	6.813	6.680	6.455
200	7.850	7.607	7.271	7.045	6.881	6.702	6.503	6.370	6.142
∞	7.531	7.295	6.970	6.750	6.590	6.414	6.217	6.084	5.850

† 乘 10^4.

* 乘 10^2.

(续)

$n\backslash m$	1% 显著性水平 $p=5$ 5	6	8	10	12	15	20	25	40
5	20.495*	20.834*	21.267*	21.53*	–	–	–	–	–
6	15.014*	15.019*	15.025*	15.029*	15.033*	15.03*	15.06*	–	–
7	2.735*	2.704*	2.665*	2.640*	2.623*	2.606*	2.590*	2.579*	–
8	1.150*	1.128*	1.099*	1.081*	1.069*	1.057*	1.044*	1.036*	–
10	48.048	46.670	44.877	43.758	42.992	42.210	41.408	40.921	–
12	31.108	30.065	28.701	27.846	27.257	26.653	26.031	25.648	25.06
14	24.016	23.145	22.001	21.279	20.781	20.268	19.736	19.408	18.90
16	20.240	19.472	18.459	17.817	17.373	16.913	16.435	16.138	15.678
18	17.929	17.228	16.302	15.713	15.304	14.878	14.435	14.159	13.727
20	16.380	15.727	14.862	14.310	13.925	13.525	13.105	12.843	12.431
25	14.107	13.529	12.759	12.265	11.918	11.555	11.172	10.930	10.547
30	12.880	12.345	11.629	11.167	10.842	10.500	10.136	9.906	9.538
35	12.115	11.607	10.926	10.486	10.174	9.845	9.494	9.271	8.911
40	11.593	11.105	10.448	10.022	9.720	9.401	9.058	8.839	8.484
50	10.928	10.465	9.841	9.434	9.144	8.836	8.504	8.290	7.940
60	10.523	10.076	9.471	9.076	8.794	8.493	8.167	7.956	7.609
70	10.251	9.814	9.223	9.835	8.559	8.263	7.941	7.732	7.386
80	10.055	9.626	9.045	8.663	8.390	8.097	7.779	7.571	7.225
100	9.793	9.374	8.806	8.432	8.164	7.876	7.561	7.355	7.009
200	9.306	8.907	8.363	8.004	7.745	7.465	7.157	6.953	6.606
∞	8.863	8.482	7.961	7.615	7.365	7.093	6.790	6.588	6.236

* 乘 10^2.

(续)

$n\backslash m$	5% 显著性水平 $p=6$								
	6	8	10	12	15	20	25	30	35
10	45.722	44.677	44.019	43.567	43.103	42.626	42.334	42.136	41.993
12	28.959	28.121	27.590	27.223	26.843	26.451	26.209	26.044	25.925
14	22.321	21.600	21.141	20.821	20.489	20.144	19.929	19.783	19.677
16	18.858	18.210	17.795	17.505	17.202	16.886	16.688	16.553	16.455
18	16.755	16.157	15.772	15.501	15.218	14.921	14.735	14.607	14.513
20	15.351	14.788	14.424	14.168	13.899	13.615	13.436	13.313	13.233
25	13.293	12.786	12.456	12.222	11.975	11.711	11.544	11.428	11.343
30	12.180	11.705	11.395	11.173	10.939	10.687	10.526	10.414	10.331
35	11.484	11.031	10.733	10.520	10.293	10.049	9.892	9.782	9.700
40	11.009	10.571	10.282	10.075	9.853	9.614	9.460	9.351	9.270
50	10.402	9.983	9.706	9.507	9.293	9.060	8.908	8.801	8.721
60	10.031	9.625	9.355	9.160	8.951	8.721	8.572	8.465	8.385
70	9.781	9.383	9.118	8.927	8.720	8.494	8.345	8.239	8.159
80	9.601	9.209	8.948	8.759	8.555	8.330	8.182	8.076	7.996
100	9.360	8.976	8.720	8.534	8.333	8.110	7.963	7.857	7.777
200	8.910	8.542	8.295	8.115	7.919	7.701	7.555	7.449	7.369
500	8.659	8.300	8.059	7.882	7.689	7.473	7.328	7.222	7.140
1000	8.579	8.223	7.983	7.808	7.616	7.400	7.255	7.149	7.067
∞	8.500	8.146	7.908	7.734	7.543	7.328	7.183	7.077	6.994

$n\backslash m$	1% 显著性水平 $p=6$								
	6	8	10	12	15	20	25	30	35
10	86.397	83.565	81.804	80.602	79.376	78.124	77.360	76.845	76.474
12	46.027	44.103	42.899	42.073	41.227	40.359	39.826	39.466	39.206
14	32.433	30.918	29.966	29.309	28.634	27.936	27.507	27.215	27.004
16	25.977	24.689	23.875	23.311	22.729	22.126	21.753	21.498	21.314
18	22.292	21.146	20.418	19.913	19.389	18.844	18.505	18.273	18.105
20	19.935	18.886	18.217	17.752	17.267	16.761	16.445	16.229	16.071
25	16.642	15.737	15.156	14.749	14.324	13.875	13.592	13.397	13.254
30	14.944	14.118	13.586	13.211	12.816	12.398	12.133	11.949	11.814
35	13.913	13.138	12.635	12.281	11.906	11.506	11.252	11.074	10.943
40	13.223	12.482	12.000	11.659	11.298	10.911	10.663	10.490	10.361
50	12.358	11.661	11.206	10.882	10.538	10.167	9.927	9.759	9.633
60	11.839	11.169	10.730	10.417	10.083	9.721	9.486	9.320	9.196
70	11.493	10.841	10.413	10.107	9.779	9.424	9.192	9.028	8.905
80	11.246	10.607	10.187	9.886	9.563	9.212	8.983	8.819	8.697
100	10.917	10.295	9.886	9.592	9.276	8.930	8.703	8.541	8.419
200	10.312	9.723	9.333	9.052	8.748	8.412	8.190	8.030	7.908
500	9.980	9.409	9.030	8.755	8.458	8.128	7.907	7.747	7.625
1000	9.874	9.308	8.933	8.661	8.365	8.037	7.817	7.657	7.534
∞	9.770	9.210	8.838	8.568	8.274	7.948	7.728	7.568	7.446

(续)

n\m	8	10	12	15	20	25	30	35
	5% 显著性水平							
	p = 7							
10	85.040	84.082	83.426	82.755	82.068	81.648	81.364	81.159
12	42.850	42.126	41.627	41.113	40.583	40.257	40.037	39.877
14	29.968	29.373	28.961	28.534	28.091	27.817	27.631	27.495
16	24.038	23.519	23.158	22.781	22.389	22.145	21.978	21.857
18	20.692	20.222	19.893	19.549	19.189	18.964	18.809	18.696
20	18.561	18.125	17.819	17.498	17.159	16.947	16.800	16.694
25	15.587	15.202	14.930	14.642	14.337	14.143	14.009	13.911
30	14.049	13.693	13.440	13.172	12.884	12.701	12.573	12.478
35	13.113	12.776	12.535	12.278	12.002	11.825	11.700	11.608
40	12.485	12.160	11.927	11.679	11.411	11.237	11.115	11.025
50	11.695	11.386	11.165	10.927	10.668	10.500	10.381	10.292
60	11.219	10.921	10.706	10.475	10.221	10.056	9.938	9.850
70	10.901	10.610	10.400	10.173	9.923	9.760	9.643	9.555
80	10.674	10.388	10.181	9.957	9.710	9.548	9.432	9.344
100	10.371	10.091	9.889	9.669	9.426	9.265	9.150	9.062
200	9.812	9.545	9.350	9.138	8.902	8.744	8.629	8.542
500	9.504	9.244	9.054	8.846	8.613	8.456	8.342	8.254
1000	9.405	9.148	8.959	8.753	8.521	8.365	8.250	8.162
∞	9.308	9.053	8.866	8.661	8.431	8.275	8.160	8.072

n\m	8	10	12	15	20	25	30	35
	1% 显著性水平							
	p = 7							
10	185.93	182.94	180.90	178.83	176.73	175.44	174.57	173.92
12	71.731	69.978	68.779	67.552	66.296	65.528	65.010	64.636
14	44.255	42.978	42.099	41.197	40.269	39.698	39.311	39.032
16	33.097	32.057	31.339	30.599	29.834	29.361	29.039	28.806
18	27.273	26.374	25.750	25.105	24.435	24.019	23.735	23.529
20	23.757	22.949	22.388	21.804	21.195	20.816	20.556	20.367
25	19.117	18.440	17.965	17.469	16.947	16.619	16.392	16.227
30	16.848	16.239	15.810	15.360	14.882	14.580	14.370	14.216
35	15.512	14.945	14.544	14.121	13.670	13.383	13.183	13.036
40	14.634	14.095	13.713	13.309	12.876	12.599	12.405	12.262
50	13.553	13.049	12.691	12.310	11.899	11.634	11.448	11.309
60	12.914	12.432	12.088	11.720	11.323	11.065	10.882	10.746
70	12.492	12.024	11.690	11.332	10.942	10.689	10.509	10.374
80	12.193	11.736	11.408	11.056	10.673	10.422	10.244	10.110
100	11.797	11.353	11.034	10.691	10.316	10.070	9.894	9.761
200	11.077	10.658	10.356	10.028	9.667	9.427	9.254	9.123
500	10.685	10.230	9.987	9.668	9.314	9.078	8.906	8.774
1000	10.561	10.160	9.869	9.553	9.202	8.966	8.795	8.663
∞	10.439	10.043	9.755	9.441	9.092	8.857	8.686	8.555

(续)

$n\backslash m$	5% 显著性水平							
	$p=8$							
	8	10	12	15	20	25	30	35
14	42.516	41.737	41.198	40.641	40.066	39.711	39.470	39.296
16	31.894	31.242	30.788	30.318	29.829	29.525	29.318	29.167
18	26.421	25.847	25.446	25.028	24.591	24.319	24.132	23.996
20	23.127	22.605	22.239	21.856	21.454	21.201	21.028	20.902
25	18.770	18.324	18.009	17.677	17.325	17.102	16.947	16.834
30	16.626	16.221	15.934	15.629	15.303	15.095	14.950	14.843
35	15.356	14.977	14.707	14.418	14.109	13.910	13.771	13.668
40	14.518	14.156	13.898	13.621	13.322	13.129	12.994	12.893
50	13.482	13.142	12.898	12.636	12.351	12.165	12.034	11.936
60	12.866	12.540	12.305	12.051	11.774	11.593	11.465	11.368
70	12.459	12.142	11.912	11.665	11.393	11.215	11.088	10.992
80	12.169	11.858	11.634	11.390	11.122	10.946	10.820	10.725
100	11.785	11.483	11.264	11.026	10.763	10.590	10.465	10.370
200	11.084	10.798	10.589	10.362	10.108	9.939	9.816	9.722
500	10.701	10.423	10.221	9.999	9.751	9.584	9.461	9.367
1000	10.579	10.304	10.104	9.884	9.637	9.470	9.348	9.254
∞	10.459	10.188	9.989	9.771	9.526	9.360	9.238	9.144

$n\backslash m$	1% 显著性水平							
	$p=8$							
	8	10	12	15	20	25	30	35
14	65.793	64.035	62.828	61.592	60.323	59.545	59.019	58.639
16	44.977	43.633	42.707	41.754	40.771	40.164	39.753	39.456
18	35.265	34.146	33.373	32.573	31.745	31.232	30.882	30.629
20	29.786	28.808	28.129	27.425	26.691	26.235	25.924	25.697
25	23.001	22.212	21.661	21.085	20.480	20.100	19.838	19.647
30	19.867	19.173	18.686	18.173	17.631	17.288	17.051	16.876
35	18.077	17.440	16.991	16.516	16.011	15.690	15.466	15.301
40	16.924	16.324	15.900	15.451	14.970	14.662	14.447	14.288
50	15.528	14.975	14.582	14.163	13.711	13.420	13.216	13.063
60	14.715	14.190	13.815	13.414	12.980	12.698	12.499	12.351
70	14.184	13.677	13.313	12.925	12.502	12.226	12.031	11.885
80	13.810	13.315	12.960	12.580	12.165	11.894	11.701	11.556
100	13.317	12.839	12.496	12.127	11.722	11.457	11.267	11.124
200	12.429	11.983	11.660	11.311	10.925	10.669	10.484	10.343
500	11.951	11.521	11.210	10.871	10.495	10.244	10.061	9.921
1000	11.800	11.375	11.067	10.732	10.359	10.109	9.927	9.787
∞	11.652	11.233	10.928	10.597	10.227	9.978	9.796	9.656

(续)

$n\backslash m$	10	12	15	20	25	30	35
	5% 显著性水平						
	$p=10$						
14	98.999	98.013	97.002	95.963	95.326	94.9	94.6
16	58.554	57.814	57.050	56.260	55.772	55.44	55.20
18	43.061	42.454	41.824	41.169	40.762	40.485	40.284
20	35.146	34.620	34.071	33.497	33.140	32.895	32.716
25	26.080	25.660	25.219	24.753	24.458	24.255	24.107
30	22.140	21.773	21.384	20.970	20.706	20.523	20.388
35	19.955	19.618	19.260	18.876	18.630	18.458	18.331
40	18.569	18.252	17.914	17.550	17.316	17.151	17.029
50	16.913	16.622	16.309	15.969	15.748	15.592	15.476
60	15.960	15.684	15.385	15.059	14.847	14.695	14.582
70	15.341	15.074	14.786	14.469	14.261	14.113	14.002
80	14.907	14.647	14.365	14.055	13.851	13.705	13.595
100	14.338	14.087	13.814	13.513	13.313	13.170	13.061
200	13.319	13.085	12.828	12.542	12.351	12.212	12.106
500	12.774	12.548	12.301	12.023	11.836	11.699	11.594
1000	12.602	12.379	12.134	11.859	11.674	11.538	11.432
∞	12.434	12.214	11.972	11.700	11.515	11.380	11.275

$n\backslash m$	10	12	15	20	25	30	35
	1% 显著性水平						
	$p=10$						
14	180.90	178.28	175.62	172.91	171.24	170	—
16	89.068	87.414	85.270	83.980	82.91	82.2	81.7
18	59.564	58.328	57.055	55.742	54.933	54.384	53.990
20	45.963	44.951	43.905	42.821	42.150	41.693	41.362
25	31.774	31.029	30.253	29.440	28.932	28.583	28.328
30	26.115	25.489	24.832	24.139	23.701	23.399	23.177
35	23.116	22.556	21.966	21.338	20.939	20.663	20.459
40	21.267	20.749	20.201	19.615	19.241	18.980	18.787
50	19.114	18.646	18.148	17.611	17.266	17.023	16.842
60	17.901	17.462	16.992	16.484	16.154	15.922	15.748
70	17.124	16.703	16.252	15.762	15.443	15.216	15.046
80	16.583	16.175	15.738	15.260	14.948	14.726	14.559
100	15.881	15.490	15.069	14.608	14.305	14.088	13.925
200	14.641	14.280	13.889	13.457	13.169	12.962	12.803
500	13.986	13.641	13.266	12.848	12.569	12.366	12.210
1000	13.780	13.441	13.070	12.658	12.381	12.179	12.023
∞	13.581	13.246	12.881	12.472	12.198	11.997	11.842

表 B.3 Bartlett-Nanda-Pillai 迹检验的分位数表

$$\Pr\left\{\frac{n+m}{m}V \geqslant x_\alpha\right\} = \alpha$$

$p = 2$

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	13	5.499	4.250	3.730	3.430	3.229	3.082	2.970	2.881	2.808	2.747	2.545	2.431
	15	5.567	4.310	3.782	3.476	3.271	3.122	3.008	2.917	2.842	2.779	2.572	2.455
	19	5.659	4.396	3.858	3.546	3.336	3.183	3.066	2.972	2.895	2.831	2.616	2.493
	23	5.718	4.453	3.911	3.595	3.383	3.228	3.109	3.013	2.935	2.869	2.650	2.524
	27	5.759	4.495	3.950	3.632	3.418	3.261	3.141	3.045	2.966	2.899	2.677	2.548
	33	5.801	4.539	3.992	3.672	3.456	3.299	3.178	3.081	3.001	2.934	2.709	2.578
	43	5.845	4.586	4.037	3.716	3.499	3.341	3.219	3.122	3.041	2.974	2.746	2.613
	63	5.891	4.635	4.086	3.764	3.547	3.389	3.266	3.169	3.088	3.020	2.791	2.657
	83	5.914	4.661	4.112	3.790	3.573	3.415	3.293	3.195	3.114	3.046	2.818	2.683
	123	5.938	4.688	4.139	3.818	3.601	3.443	3.321	3.223	3.143	3.075	2.847	2.713
	243	5.962	4.715	4.168	3.846	3.630	3.472	3.351	3.254	3.174	3.106	2.880	2.748
	∞	5.991	4.744	4.197	3.877	3.661	3.504	3.384	3.287	3.208	3.141	2.918	2.788
.01	13	7.499	5.409	4.570	4.094	3.780	3.555	3.383	3.248	3.138	3.047	2.751	2.587
	15	7.710	5.539	4.671	4.180	3.857	3.625	3.448	3.309	3.196	3.101	2.795	2.625
	19	8.007	5.732	4.824	4.312	3.976	3.734	3.550	3.405	3.287	3.188	2.867	2.686
	23	8.206	5.868	4.935	4.409	4.064	3.815	3.627	3.478	3.356	3.255	2.923	2.735
	27	8.349	5.970	5.019	4.483	4.131	3.878	3.686	3.534	3.410	3.307	2.968	2.775
	33	5.500	6.080	5.111	4.566	4.207	3.950	3.754	3.600	3.473	3.368	3.021	2.823
	43	8.660	6.201	5.214	4.659	4.294	4.032	3.833	3.675	3.547	3.439	3.085	2.881
	63	8.831	6.333	5.329	4.764	4.393	4.127	3.925	3.765	3.634	3.525	3.163	2.955
	83	8.920	6.404	5.392	4.823	4.449	4.181	3.977	3.815	3.684	3.574	3.210	3.000
	123	9.012	6.478	5.459	5.885	4.508	4.238	4.033	3.871	3.739	6.628	3.263	3.052
	243	9.108	6.556	5.529	4.951	4.572	4.301	4.095	3.932	3.800	3.689	3.323	3.113
	∞	9.210	6.638	5.604	5.023	4.642	4.369	4.163	4.000	3.867	3.757	3.393	3.185

(续)

α	$n \backslash m$	1	2	3	4	5	6	7	8	9	10	15	20
							$p = 3$						
.05	14	6.989	5.595	5.019	4.684	4.458	4.293	4.165	4.063	3.979	3.908	3.672	3.537
	16	7.095	5.673	5.082	4.738	4.507	4.338	4.207	4.103	4.017	3.944	3.702	3.563
	20	7.243	5.787	5.177	4.822	4.583	4.409	4.274	4.166	4.077	4.002	3.751	3.606
	24	7.341	5.866	5.245	4.883	4.639	4.461	4.323	4.213	4.122	4.046	3.790	3.640
	28	7.410	5.925	5.295	4.929	4.682	4.501	4.362	4.250	4.158	4.081	3.821	3.668
	34	7.482	5.987	5.351	4.980	4.730	4.547	4.406	4.293	4.200	4.121	3.857	3.702
	44	7.559	6.055	5.412	5.037	4.784	4.599	4.457	4.342	4.248	4.169	3.901	3.743
	64	7.639	6.129	5.480	5.101	4.846	4.660	4.516	4.400	4.305	4.225	3.955	3.795
	84	7.681	6.168	5.517	5.137	4.880	4.693	4.549	4.433	4.338	4.257	3.986	3.826
	124	7.724	6.209	5.556	5.174	4.917	4.730	4.585	4.469	4.374	4.293	4.022	3.862
	244	7.768	6.251	5.597	5.214	4.957	4.769	4.624	4.508	4.413	4.333	4.063	3.904
	∞	7.815	6.296	5.640	5.257	4.999	4.812	4.667	4.552	4.457	4.377	4.110	3.954
.01	14	8.971	6.855	5.970	5.457	5.112	4.862	4.669	4.516	4.390	4.285	3.939	3.743
	16	9.245	7.006	6.083	5.551	5.195	4.937	4.738	4.581	4.451	4.343	3.986	3.783
	20	9.639	7.236	6.258	5.698	5.326	5.056	4.849	4.684	4.549	4.436	4.063	3.850
	24	9.910	7.403	6.387	5.808	5.424	5.146	4.933	4.764	4.625	4.509	4.124	3.903
	28	10.106	7.528	6.486	5.893	5.501	5.217	4.999	4.827	4.685	4.567	4.174	3.948
	34	10.317	7.667	6.598	5.990	5.588	5.298	5.076	4.900	4.756	4.635	4.233	4.001
	44	10.545	7.821	6.724	6.101	5.690	5.393	5.167	4.986	4.839	4.715	4.305	4.067
	64	10.790	7.994	6.867	6.230	5.809	5.505	5.274	5.090	4.939	4.813	4.394	4.151
	84	10.920	8.088	6.947	6.301	5.876	5.569	5.335	5.150	4.998	4.871	4.448	4.203
	124	11.056	8.188	7.032	6.379	5.948	5.639	5.403	5.216	5.063	4.935	4.510	4.263
	244	11.196	8.294	7.124	6.463	6.028	5.716	5.478	5.290	5.136	5.007	4.581	4.334
	∞	11.345	8.406	7.222	6.554	6.116	5.801	5.562	5.372	5.218	5.089	4.664	4.419

(续)

α	$n \backslash m$	1	2	3	4	5	6	7	8	9	10	15	20
							$p=4$						
.05	15	8.331	6.859	6.245	5.885	5.642	5.462	5.323	5.212	5.119	5.041	4.779	4.627
	17	8.472	6.952	6.318	5.947	5.696	5.512	5.369	5.255	5.160	5.080	4.811	4.654
	21	8.671	7.091	6.429	6.043	5.782	5.591	5.443	5.324	5.225	5.143	4.864	4.701
	25	8.805	7.190	6.510	6.114	5.846	5.650	5.498	5.377	5.276	5.191	4.906	4.738
	29	8.901	7.263	6.571	6.168	5.896	5.696	5.542	5.418	5.316	5.230	4.939	4.768
	35	9.004	7.343	6.640	6.229	5.952	5.749	5.593	5.467	5.363	5.275	4.980	4.805
	45	9.113	7.431	6.716	6.298	6.017	5.811	5.652	5.524	5.418	5.329	5.029	4.851
	65	9.229	7.528	6.802	6.378	6.092	5.883	5.721	5.592	5.485	5.395	5.090	4.910
	85	9.291	7.580	6.849	6.421	9.134	5.923	5.761	5.631	5.523	5.432	5.127	4.945
	125	9.354	7.635	6.899	6.469	6.179	5.968	5.804	5.674	5.566	5.475	5.168	4.987
	245	9.419	7.693	6.952	6.519	6.228	6.016	5.852	5.721	5.613	5.522	5.216	5.035
	∞	9.488	7.754	7.009	6.574	6.282	6.069	5.905	5.774	5.667	5.576	5.272	5.094
.01	15	10.293	8.188	7.276	6.737	6.737	6.105	5.898	5.731	5.594	5.479	5.095	4.874
	17	10.619	8.360	7.401	6.840	6.462	6.184	5.971	5.799	5.658	5.539	5.144	4.916
	21	11.095	8.625	7.598	7.003	6.604	6.313	6.089	5.909	5.762	5.638	5.225	4.987
	25	11.428	8.818	7.744	7.126	6.712	6.411	6.180	5.995	5.844	5.716	5.290	5.044
	29	11.672	8.966	7.858	7.222	6.798	6.490	6.253	6.604	5.909	5.779	5.344	5.091
	35	11.938	9.131	7.987	7.332	6.897	6.581	6.338	6.145	5.986	5.853	5.408	5.149
	45	12.228	9.318	8.135	7.460	7.012	6.688	6.439	6.241	6.079	5.942	5.487	5.221
	65	12.545	9.529	8.306	7.610	7.149	6.816	6.561	6.358	6.192	6.052	5.586	5.314
	85	12.715	9.645	8.402	7.695	7.227	6.890	6.632	6.426	6.258	6.117	5.646	5.371
	125	12.893	9.769	8.505	7.787	7.313	6.971	6.710	6.503	6.333	6.190	5.715	5.439
	245	13.080	9.902	8.617	7.889	7.408	7.062	6.798	6.588	6.417	6.273	5.796	5.519
	∞	13.277	10.045	8.739	8.000	7.513	7.163	6.897	6.686	6.513	6.369	5.892	5.616

(续)

$p = 5$

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
	16	9.589	8.071	7.430	7.052	6.795	6.605	6.457	6.338	6.239	6.155	5.873	5.706
	18	9.761	8.179	7.512	7.120	6.854	6.659	6.506	6.384	6.282	6.196	5.906	5.735
	22	10.007	8.340	7.639	7.228	6.949	6.745	6.586	6.458	6.352	6.263	5.961	5.784
	26	10.176	8.457	7.732	7.308	7.021	6.810	6.647	6.516	6.407	6.316	6.006	5.823
	30	10.298	8.544	7.803	7.370	7.077	6.862	6.696	6.562	6.451	6.358	6.042	5.856
.05	36	10.429	8.641	7.883	7.440	7.141	6.922	6.752	6.616	6.503	6.408	6.086	5.896
	46	10.571	8.748	7.974	7.521	7.216	6.992	6.819	6.680	6.565	6.468	6.140	5.945
	66	10.724	8.868	8.077	7.615	7.303	7.075	6.899	6.757	6.640	6.541	6.208	6.009
	86	10.805	8.933	8.134	7.667	7.353	7.122	6.944	6.802	6.684	6.584	6.248	6.049
	126	10.890	9.002	8.195	7.724	7.407	7.174	6.995	6.851	6.733	6.633	6.295	6.095
	246	10.978	9.076	8.261	7.786	7.466	7.232	7.052	6.907	6.788	6.688	6.350	6.150
	∞	11.071	9.154	8.332	7.853	7.531	7.296	7.115	6.970	6.851	6.750	6.414	6.217
	16	11.534	9.451	8.521	7.966	7.587	7.306	7.088	6.912	6.767	6.644	6.230	5.989
	18	11.902	9.642	8.658	8.077	7.682	7.391	7.165	6.983	6.833	6.707	6.281	6.033
	22	12.449	9.939	8.876	8.255	7.835	7.528	7.291	7.100	6.943	6.810	6.366	6.106
	26	12.837	10.159	9.040	8.390	7.954	7.635	7.389	7.192	7.030	6.893	6.434	6.166
	30	13.125	10.328	9.168	8.497	8.048	7.720	7.468	7.266	7.100	6.960	6.491	6.216
.01	36	13.442	10.518	9.314	8.621	8.158	7.820	7.561	7.354	7.183	7.040	6.559	6.277
	46	13.790	10.735	9.483	8.765	8.287	7.939	7.673	7.460	7.284	7.137	6.644	6.355
	66	14.176	10.984	9.681	8.936	8.442	8.083	7.808	7.589	7.409	7.258	6.752	6.455
	86	14.385	11.122	9.793	9.034	8.531	8.167	7.888	7.666	7.483	7.330	6.818	6.518
	126	14.606	11.270	9.914	9.142	8.630	8.260	7.977	7.752	7.567	7.412	6.895	6.592
	246	14.839	11.431	10.047	9.260	8.740	8.364	8.077	7.849	7.663	7.506	6.985	6.681
	∞	15.086	11.605	10.193	9.392	8.863	8.482	8.192	7.961	7.773	7.615	7.093	6.790

(续)

$p = 6$

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	17	10.794	9.247	8.585	8.193	7.926	7.728	7.573	7.448	7.344	7.256	6.956	6.778
	19	10.993	9.367	8.676	8.268	7.990	7.785	7.625	7.496	7.389	7.299	6.990	6.808
	23	11.282	9.550	8.817	8.386	8.093	7.878	7.711	7.576	7.464	7.369	7.048	6.858
	27	11.483	9.684	8.922	8.475	8.172	7.950	7.777	7.638	7.523	7.425	7.095	6.899
	31	11.630	9.784	9.003	8.545	8.234	8.007	7.830	7.688	7.570	7.471	7.134	6.934
	37	11.790	9.897	9.094	8.624	8.306	8.073	7.892	7.747	7.627	7.525	7.181	6.976
	47	11.964	10.024	9.199	8.716	8.390	8.151	7.966	7.817	7.694	7.590	7.239	7.029
	67	12.154	10.166	9.319	8.824	8.490	8.245	8.056	7.903	7.778	7.672	7.312	7.099
	87	12.255	10.245	9.387	8.885	8.547	8.299	8.108	7.954	7.827	7.720	7.357	7.142
	127	12.362	10.328	9.459	8.951	8.609	8.359	8.165	8.010	7.882	7.774	7.409	7.193
	247	12.474	10.417	9.538	9.024	8.678	8.425	8.230	8.074	7.945	7.836	7.470	7.254
	∞	12.592	10.513	9.623	9.104	8.755	8.500	8.303	8.146	8.017	7.908	7.543	7.328
.01	17	12.722	10.664	9.724	9.157	8.767	8.478	8.252	8.069	7.917	7.788	7.351	7.093
	19	13.126	10.874	9.873	8.277	8.869	8.567	8.332	8.143	7.986	7.853	7.403	7.137
	23	13.736	11.202	10.111	9.469	9.034	8.714	8.465	8.266	8.100	7.961	7.490	7.213
	27	14.173	11.446	10.292	9.617	9.162	8.828	8.570	8.363	8.192	8.048	7.561	7.275
	31	14.501	11.635	10.433	9.734	9.264	8.921	8.655	8.442	8.267	8.119	7.621	7.328
	37	14.865	11.850	10.596	9.871	9.384	9.030	8.756	8.537	8.356	8.204	7.693	7.392
	47	15.270	12.097	10.787	10.032	9.527	9.160	8.878	8.652	8.466	8.309	7.783	7.474
	67	15.723	12.382	11.011	10.224	9.700	9.319	9.027	8.794	8.602	8.440	7.899	7.581
	87	15.970	12.542	11.138	10.335	9.800	9.413	9.115	8.878	8.683	8.520	7.971	7.649
	127	16.233	12.715	11.278	10.457	9.912	9.517	9.215	8.974	8.776	8.610	8.055	7.729
	247	16.513	12.903	11.432	10.593	10.037	9.635	9.328	9.084	8.883	8.715	8.154	7.827
	∞	16.812	13.108	11.602	10.745	10.178	9.770	9.458	9.210	9.008	8.838	8.274	7.948

(续)

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
							$p=7$						
.05	18	11.961	10.396	9.719	9.316	9.040	8.835	8.675	8.545	8.437	8.345	8.031	7.843
	20	12.184	10.528	9.817	9.396	9.109	8.896	8.730	8.596	8.484	8.390	8.067	7.874
	24	12.513	10.731	9.972	9.525	9.220	8.996	8.821	8.680	8.563	8.464	8.127	7.926
	28	12.744	10.880	10.088	9.622	9.306	9.073	8.892	8.746	8.626	8.523	8.176	7.969
	32	12.915	10.994	10.178	9.699	9.374	9.135	8.950	8.800	8.676	8.572	8.216	8.005
	38	13.102	11.123	10.281	9.787	9.453	9.208	9.017	8.864	8.737	8.630	8.266	8.049
	48	13.308	11.267	10.399	9.890	9.547	9.294	9.098	8.941	8.811	8.701	8.328	8.106
	68	13.534	11.433	10.537	10.012	9.658	9.398	9.197	9.036	8.902	8.789	8.407	8.180
	88	13.657	11.524	10.614	10.082	9.722	9.459	9.255	9.092	8.956	8.842	8.456	8.226
	128	13.786	11.623	10.698	10.158	9.793	9.526	9.320	9.155	9.018	8.903	8.513	8.281
	248	13.923	11.728	10.790	10.242	9.872	9.602	9.394	9.226	9.089	8.972	8.580	8.348
	∞	14.067	11.842	10.890	10.334	9.960	9.687	9.477	9.309	9.170	9.053	8.661	8.431
.01	18	13.874	11.841	10.895	10.321	9.923	9.627	9.395	9.206	9.049	8.915	8.460	8.188
	20	14.310	12.069	11.056	10.448	10.031	9.721	9.479	9.283	9.121	8.982	8.512	8.233
	24	14.974	12.426	11.314	10.655	10.207	9.876	9.619	9.412	9.240	9.095	8.602	8.310
	28	15.456	12.694	11.510	10.815	10.344	9.999	9.731	9.515	9.337	9.186	8.676	8.374
	32	15.822	12.902	11.665	10.943	10.455	10.098	9.821	9.599	9.416	9.261	8.738	8.429
	38	16.230	13.141	11.845	11.092	10.586	10.215	9.930	9.700	9.511	9.351	8.814	8.496
	48	16.688	13.416	12.056	11.269	10.742	10.357	10.061	9.824	9.628	9.463	8.909	8.582
	68	17.206	13.737	12.306	11.482	10.932	10.532	10.224	9.978	9.776	9.605	9.033	8.696
	88	17.491	13.919	12.449	11.605	11.043	10.635	10.321	10.070	9.865	9.691	9.110	8.768
	128	17.796	14.116	12.607	11.743	11.168	10.751	10.431	10.176	9.967	9.790	9.201	8.854
	248	18.124	14.333	12.782	11.897	11.308	10.883	10.557	10.298	10.085	9.906	9.309	8.960
	∞	18.475	14.571	12.977	12.070	11.468	11.034	10.703	10.439	10.223	10.043	9.441	9.092

(续)

$p=8$													
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	19	13.101	11.524	10.835	10.423	10.141	9.930	9.766	9.632	9.521	8.426	9.100	8.904
	21	13.346	11.667	10.941	10.509	10.214	9.995	9.824	9.685	9.570	9.472	9.136	8.935
	25	13.710	11.889	11.109	10.647	10.333	1.101	9.920	9.774	9.652	9.550	9.198	8.988
	29	13.970	12.054	11.235	10.753	10.425	10.184	9.996	9.844	9.718	9.612	9.249	9.032
	33	14.163	12.180	11.334	10.837	10.499	10.250	10.057	9.902	9.772	8.663	9.292	9.070
	39	14.377	12.323	11.448	10.934	10.585	10.329	10.130	9.970	9.837	9.725	9.344	9.116
	49	14.614	12.487	11.580	11.048	10.688	10.423	10.218	10.053	9.917	9.801	9.409	9.176
	69	14.877	12.674	11.734	11.183	10.811	10.538	10.326	10.156	10.016	9.897	9.494	9.254
	89	15.021	12.779	11.822	11.261	10.883	10.605	10.390	10.218	10.075	9.955	9.547	9.304
	129	15.173	12.892	11.918	11.347	10.962	10.680	10.462	10.288	10.143	10.021	9.609	9.363
	249	15.335	13.015	12.023	11.442	11.051	10.765	10.544	10.367	10.221	10.098	9.682	9.436
	∞	15.507	13.148	12.138	11.548	11.152	10.862	10.638	10.459	10.312	10.188	9.771	9.526
.01	19	14.999	12.992	12.043	11.463	11.060	10.758	10.521	10.328	10.167	10.030	9.558	9.275
	21	15.463	13.235	12.215	11.598	11.174	10.857	10.610	10.409	10.241	10.099	9.612	9.321
	25	16.177	13.620	12.491	11.819	11.360	11.021	10.757	10.543	10.366	10.216	9.704	9.399
	29	16.700	13.910	12.703	11.991	11.507	11.151	10.874	10.651	10.467	10.310	9.780	9.465
	33	17.100	14.137	12.871	12.128	11.626	11.256	10.970	10.740	10.550	10.389	9.844	9.521
	39	17.549	14.398	13.067	12.290	11.766	11.383	11.086	10.848	10.651	10.485	9.924	9.591
	49	18.058	14.702	13.297	12.482	11.935	11.536	11.277	10.980	10.776	10.604	10.024	9.681
	69	18.640	15.058	13.573	12.716	12.143	11.725	11.403	11.146	10.934	10.756	10.155	9.801
	89	18.962	15.261	13.733	12.853	12.265	11.838	11.509	11.247	11.030	10.849	10.238	9.877
	129	19.310	15.484	13.909	13.005	12.403	11.965	11.630	11.362	11.142	10.956	10.335	9.970
	249	19.684	15.729	14.406	13.177	12.559	12.112	11.769	11.495	11.271	11.083	10.453	10.083
	∞	20.090	16.000	14.327	13.371	12.738	12.280	11.930	11.652	11.424	11.233	10.597	10.277

(续)

$p = 10$

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	21	15.322	13.732	13.027	12.603	12.311	12.093	11.922	11.782	11.666	11.566	11.222	11.013
	23	15.604	13.897	13.147	12.700	12.393	12.164	11.985	11.840	11.719	11.616	11.260	11.045
	27	16.033	14.154	13.340	12.857	12.526	12.282	12.091	11.937	11.808	11.699	11.325	11.100
	31	16.344	14.347	13.487	12.978	12.631	12.375	12.176	12.015	11.881	11.768	11.380	11.146
	35	16.580	14.497	13.603	13.075	12.716	12.451	12.245	12.079	11.941	11.824	11.426	11.186
	41	16.843	14.669	13.737	13.188	12.816	12.541	12.328	12.156	12.014	11.893	11.482	11.236
	51	17.140	14.868	13.895	13.323	12.936	12.651	12.430	12.251	12.104	11.979	11.555	11.301
	71	17.476	15.100	14.083	13.486	13.083	12.786	12.556	12.371	12.218	12.089	11.650	11.388
	91	17.662	15.231	14.191	13.581	13.169	12.866	12.632	12.443	12.288	12.156	11.710	11.443
	131	17.861	15.375	14.310	13.687	13.266	12.957	12.718	12.526	12.368	12.234	11.781	11.511
	251	18.076	15.532	14.443	13.806	13.376	13.061	12.818	12.622	12.461	12.325	11.867	11.594
	∞	18.307	15.705	14.591	13.940	13.501	13.180	12.933	12.735	12.572	12.434	11.972	11.700
.01	21	17.197	15.234	14.284	13.698	13.288	12.980	12.736	12.537	12.371	12.228	11.733	11.432
	23	17.707	15.507	14.476	13.849	13.413	13.088	12.832	12.624	12.449	12.301	11.789	11.478
	27	18.505	15.941	14.788	14.096	13.621	13.268	12.993	12.769	12.584	12.426	11.885	11.559
	31	19.101	16.273	15.029	14.290	13.786	13.413	13.123	12.888	12.693	12.528	11.965	11.628
	35	19.562	16.535	15.222	14.447	13.920	13.531	13.230	12.986	12.785	12.614	12.034	11.687
	41	20.088	16.839	15.448	14.632	14.080	13.674	13.359	13.106	12.897	12.720	12.119	11.761
	51	20.692	17.196	15.718	14.855	14.274	13.848	13.519	13.255	13.037	12.852	12.229	11.858
	71	21.394	17.623	16.045	15.130	14.516	14.067	13.722	13.445	13.216	13.024	12.374	11.989
	91	21.790	17.868	16.236	15.292	14.660	14.199	13.845	13.561	13.327	13.130	12.466	12.074
	131	22.221	18.141	16.450	15.475	14.824	14.350	13.986	13.695	13.456	13.254	12.577	12.177
	251	22.692	18.444	16.690	15.683	15.012	14.525	14.152	13.853	13.608	13.402	12.712	12.307
	∞	23.209	18.783	16.964	15.923	15.231	14.730	14.346	14.041	13.791	13.581	12.881	12.472

表 B.4 Roy 最大根检验的分位数表

$$\Pr\left\{\frac{m+n}{m}R \geqslant x_\alpha\right\} = \alpha$$

$p = 2$

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	13	5.499	3.736	3.011	2.605	2.342	2.157	2.018	1.910	1.823	1.752	1.527	1.407
	15	5.567	3.807	3.078	2.668	2.401	2.211	2.069	1.959	1.869	1.796	1.562	1.436
	19	5.659	3.905	3.173	2.759	2.487	2.293	2.148	2.033	1.940	1.864	1.618	1.484
	23	5.718	3.971	3.239	2.822	2.548	2.352	2.204	2.087	1.993	1.915	1.661	1.521
	27	5.759	4.018	3.286	2.868	2.593	2.396	2.247	2.129	2.033	1.954	1.696	1.552
	33	5.801	4.068	3.336	2.918	2.643	2.445	2.294	2.175	2.079	1.998	1.736	1.588
	43	5.845	4.120	3.391	2.973	2.697	2.498	2.347	2.228	2.131	2.049	1.783	1.631
	63	5.891	4.176	3.449	3.032	2.757	2.558	2.407	2.288	2.190	2.109	1.840	1.686
	83	5.914	4.205	3.480	3.064	2.789	2.591	2.440	2.321	2.223	2.142	1.873	1.718
	123	5.938	4.235	3.512	3.097	2.823	2.626	2.476	2.356	2.259	2.178	1.909	1.755
	243	5.962	4.265	3.545	3.132	2.859	2.663	2.513	2.395	2.298	2.217	1.951	1.797
	∞	5.991	4.297	3.580	3.169	2.897	2.702	2.554	2.436	2.340	2.261	1.998	1.847
.01	13	7.499	4.675	4.610	3.040	2.681	2.432	2.249	2.109	1.997	1.907	1.625	1.478
	15	7.710	4.834	3.742	3.154	2.782	2.523	2.333	2.186	2.069	1.973	1.676	1.519
	19	8.007	5.064	3.937	3.325	2.936	2.664	2.463	2.307	2.182	2.080	1.758	1.587
	23	8.206	5.223	4.074	3.448	3.048	2.768	2.559	2.397	2.268	2.161	1.823	1.641
	27	8.349	5.339	4.176	3.540	3.133	2.847	2.634	2.468	2.335	2.225	1.876	1.686
	33	8.500	5.465	4.287	3.642	3.228	2.936	2.718	2.548	2.412	2.299	1.938	1.740
	43	8.660	5.600	4.409	3.755	3.334	3.037	2.815	2.641	2.501	2.386	2.013	1.807
	63	8.831	5.747	4.543	3.881	3.454	3.153	2.926	2.749	2.607	2.488	2.105	1.891
	83	8.920	5.825	4.616	3.950	3.520	3.217	2.989	2.810	2.666	2.547	2.160	1.943
	123	9.012	5.906	4.692	4.022	3.591	3.285	3.056	2.877	2.732	2.612	2.222	2.002
	243	9.108	5.991	4.772	4.100	3.666	3.360	3.130	2.950	2.804	2.683	2.292	2.072
	∞	9.210	6.080	4.856	4.182	3.747	3.440	3.209	3.029	2.884	2.763	2.373	2.154

(续)

							$p=3$						
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	14	6.989	4.517	3.544	3.010	2.669	2.430	2.254	2.117	2.008	1.919	1.639	1.491
	16	7.095	4.617	3.634	3.092	2.745	2.501	2.319	2.178	2.065	1.973	1.682	1.526
	20	7.243	4.760	3.767	3.215	2.859	2.608	2.420	2.274	2.156	2.059	1.751	1.585
	24	7.341	4.858	3.859	3.302	2.942	2.686	2.495	2.345	2.224	2.124	1.805	1.631
	28	7.410	4.929	3.927	3.367	3.004	2.746	2.552	2.400	2.277	2.176	1.849	1.669
	34	7.482	5.004	4.001	3.439	3.073	2.812	2.616	2.462	2.338	2.234	1.901	1.715
	44	7.559	5.086	4.081	3.517	3.149	2.887	2.689	2.534	2.408	2.303	1.962	1.771
	64	7.639	5.173	4.169	3.604	3.325	2.972	2.773	2.616	2.489	2.383	2.038	1.842
	84	7.681	5.220	4.216	3.651	3.282	3.019	2.820	2.663	2.535	2.429	2.082	1.885
	124	7.724	5.268	4.265	3.701	3.332	3.069	2.870	2.713	2.586	2.479	2.132	1.934
	244	7.768	5.318	4.317	3.754	3.386	3.123	2.924	2.768	2.641	2.535	2.189	1.991
	∞	7.815	5.370	4.371	3.810	3.443	3.181	2.983	2.828	2.701	2.596	2.253	2.059
.01	14	8.971	5.416	4.106	3.412	2.978	2.680	4.462	2.295	2.163	2.055	1.724	1.552
	16	9.245	5.613	4.265	3.548	3.098	2.787	2.559	2.384	2.245	2.132	1.782	1.598
	20	9.639	5.905	4.507	3.757	3.284	2.956	2.714	2.527	2.378	2.257	1.877	1.676
	24	9.910	6.111	4.681	3.910	3.422	3.082	2.831	2.636	2.481	2.354	1.954	1.740
	28	10.106	6.264	4.811	4.026	3.528	3.180	2.922	2.722	2.562	2.431	2.016	1.792
	34	10.317	6.431	4.955	4.156	3.647	3.292	3.027	2.821	2.657	2.521	2.091	1.857
	44	10.545	6.614	5.116	4.303	3.784	3.420	3.148	2.938	2.768	2.628	2.182	1.937
	64	10.790	6.815	5.296	4.469	3.940	3.568	3.291	3.075	2.901	2.757	2.295	2.040
	84	10.920	6.923	5.393	4.560	4.027	3.652	3.372	3.153	2.977	2.832	2.363	2.103
	124	11.056	7.037	2.497	4.658	4.120	3.742	3.460	3.239	3.062	2.915	2.441	2.177
	244	11.196	7.157	5.608	4.763	4.221	3.841	3.556	3.334	3.155	3.008	2.530	2.264
	∞	11.345	7.284	5.725	4.875	4.331	3.948	3.663	3.440	3.260	3.112	2.634	2.369

(续)

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
							$p=4$						
.05	15	8.331	5.211	4.013	3.365	2.955	2.670	2.459	2.297	2.168	2.063	1.736	1.564
	17	8.472	5.336	4.123	3.464	3.045	2.752	2.536	2.368	2.235	2.126	1.784	1.603
	21	8.671	5.517	4.287	3.613	3.182	2.880	2.655	2.481	2.341	2.227	1.864	1.670
	25	8.805	5.644	4.403	3.721	3.283	2.975	2.745	2.566	2.422	2.304	1.928	1.724
	29	8.901	5.736	4.490	3.802	3.360	3.048	2.814	2.632	2.486	2.365	1.979	1.769
	35	9.004	5.837	4.585	3.893	3.446	3.130	2.893	2.708	2.559	2.436	2.040	1.822
	45	9.113	5.946	4.690	3.993	3.543	3.224	2.984	2.796	2.645	2.520	2.114	1.889
	65	9.229	6.065	4.806	4.106	3.653	3.332	3.090	2.900	2.746	2.619	2.206	1.974
	85	9.291	6.128	4.869	4.168	3.714	3.392	3.149	2.959	2.804	2.677	2.261	2.026
	125	9.354	6.195	4.935	4.234	3.779	3.457	3.214	3.023	2.868	2.740	2.322	2.087
	245	9.419	6.265	5.005	4.304	3.850	3.527	3.284	3.093	2.939	2.811	2.393	2.157
	∞	9.488	6.338	5.080	4.380	3.926	3.603	3.361	3.171	3.017	2.890	2.475	2.242
.01	15	10.293	6.080	4.549	3.744	3.244	2.901	2.651	2.461	2.310	2.188	1.812	1.618
	17	10.619	6.308	4.731	3.898	3.378	3.021	2.760	2.559	2.401	2.273	1.875	1.668
	21	11.095	6.650	5.010	4.137	3.589	3.211	2.933	2.720	2.550	2.412	1.981	1.754
	25	11.428	6.896	5.213	4.315	3.748	3.356	3.067	2.844	2.666	2.521	2.066	1.825
	29	11.672	7.080	5.368	4.451	3.872	3.469	3.172	2.942	2.759	2.609	2.137	1.884
	35	11.938	7.284	5.542	4.606	4.013	3.600	3.294	3.057	2.868	2.713	2.221	1.956
	45	12.228	7.510	5.737	4.782	4.175	3.752	3.437	3.193	2.998	2.837	2.326	2.048
	65	12.545	7.762	5.958	4.984	4.364	3.930	3.607	3.356	3.155	2.989	2.458	2.166
	85	12.715	7.899	6.080	5.096	4.470	4.031	3.704	3.450	3.246	3.078	2.538	2.240
	125	12.893	8.045	6.210	5.218	4.585	4.142	3.812	3.555	3.348	3.178	2.630	2.326
	245	13.080	8.199	6.350	6.349	4.710	4.263	3.930	3.671	3.462	3.290	2.737	2.430
	∞	13.277	8.363	6.500	5.491	4.848	4.397	4.062	3.801	3.591	3.418	2.863	2.555

(续)

α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
							$p=5$						
.05	16	9.589	5.856	4.448	3.694	3.218	2.890	2.648	2.463	2.316	2.196	1.824	1.630
	18	9.761	6.002	4.575	3.806	3.320	2.982	2.734	2.542	2.390	2.265	1.878	1.674
	22	10.007	6.218	4.765	3.978	3.477	3.128	2.869	2.669	2.509	2.378	1.967	1.747
	26	10.176	6.370	4.903	4.104	3.593	3.237	2.971	2.765	2.601	2.465	2.037	1.807
	30	10.298	6.483	5.006	4.200	3.683	3.321	3.051	2.842	2.674	2.535	2.095	1.857
	36	10.429	6.606	5.121	4.307	3.785	3.418	3.144	2.930	2.759	2.617	2.165	1.918
	46	10.571	6.742	5.249	4.428	3.901	3.529	3.251	3.034	2.859	2.714	2.250	1.994
	66	10.724	6.892	5.392	4.566	4.034	3.658	3.377	3.156	2.979	2.832	2.357	2.092
	86	10.805	6.973	5.471	4.643	4.108	3.731	3.448	3.227	3.048	2.900	2.421	2.152
	126	10.890	7.058	5.554	4.724	4.189	3.810	3.527	3.304	3.125	2.976	2.494	2.223
	246	10.978	7.148	5.643	4.812	4.276	3.897	3.613	3.390	3.210	3.061	2.578	2.306
	∞	11.071	7.244	5.738	4.907	4.371	3.992	3.708	3.485	3.306	3.157	2.676	2.407
.01	16	11.534	6.701	4.963	4.055	3.492	3.108	2.829	2.616	2.448	2.312	1.895	1.680
	18	11.902	6.954	5.163	4.223	3.638	3.238	2.945	2.722	2.546	2.403	1.962	1.733
	22	12.449	7.340	5.473	4.487	3.870	3.446	3.135	2.896	2.707	2.554	2.076	1.825
	26	12.837	7.620	5.703	4.685	4.047	3.606	3.282	3.033	2.835	2.673	2.169	1.902
	30	13.125	7.832	5.880	4.840	4.186	3.733	3.400	3.142	2.938	2.770	2.246	1.966
	36	13.442	8.069	6.079	5.016	4.346	3.881	3.537	3.271	3.060	2.886	2.339	2.046
	46	13.790	8.335	6.306	5.219	4.532	4.053	3.699	3.425	3.206	3.026	2.456	2.147
	66	14.176	8.635	6.566	5.454	4.750	4.259	3.894	3.611	3.385	3.198	2.604	2.279
	86	14.385	8.800	6.711	5.587	4.874	4.377	4.007	3.720	3.490	3.301	2.695	2.362
	126	14.606	8.977	6.867	5.731	5.010	4.506	4.132	3.841	3.608	3.416	2.800	2.461
	246	14.839	9.165	7.036	5.888	5.159	4.650	4.272	3.978	3.741	3.547	2.924	2.579
	∞	15.086	9.367	7.218	6.060	5.324	4.810	4.428	4.132	3.893	3.697	3.070	2.724

(续)

$p=6$													
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	17	10.794	6.470	4.861	4.005	3.468	3.098	2.827	2.620	2.455	2.322	1.908	1.693
	19	10.993	6.634	5.001	4.128	3.579	3.199	2.920	2.705	2.535	2.396	1.965	1.740
	23	11.282	6.880	5.216	4.320	3.753	3.359	3.068	2.844	2.665	2.519	2.062	1.819
	27	11.483	7.056	5.372	4.462	3.884	3.481	3.182	2.951	2.767	2.615	2.139	1.884
	31	11.630	7.188	5.491	4.571	3.985	3.576	3.272	3.036	2.848	2.693	2.203	1.939
	37	11.790	7.334	5.625	4.695	4.101	3.686	3.376	3.136	2.943	2.785	2.280	2.006
	47	11.964	7.495	5.774	4.836	4.235	3.813	3.498	3.253	3.057	2.894	2.375	2.090
	67	12.154	7.675	5.944	4.997	4.390	3.963	3.643	3.394	3.194	3.028	2.495	2.200
	87	12.255	7.774	6.038	5.088	4.477	4.048	3.727	3.476	3.274	3.107	2.568	2.268
	127	12.362	7.878	6.138	5.185	4.573	4.141	3.818	3.566	3.363	3.195	2.652	2.348
	247	12.474	7.989	6.246	5.291	4.676	4.244	3.920	3.667	3.463	3.294	2.749	2.444
	∞	12.592	8.107	6.362	5.405	4.790	4.357	4.033	3.780	3.576	3.408	2.864	2.561
.01	17	12.722	7.296	5.360	4.352	3.730	3.306	2.998	2.764	2.580	2.431	1.974	1.739
	19	13.126	7.570	5.574	4.531	3.885	3.444	3.122	2.877	2.683	2.526	2.044	1.795
	23	13.736	7.992	5.912	4.817	4.135	3.667	3.325	3.063	2.855	2.687	2.165	1.892
	27	14.173	8.303	6.164	5.034	4.328	3.841	3.484	3.210	2.993	2.816	2.264	1.974
	31	14.501	8.541	6.360	5.204	4.480	3.980	3.612	3.329	3.104	2.921	2.347	2.043
	37	14.865	8.808	6.583	5.400	4.657	4.142	3.763	3.470	3.237	3.047	2.448	2.128
	47	15.270	9.112	6.839	5.628	4.864	4.334	3.943	3.640	3.399	3.201	2.576	2.238
	67	15.723	9.458	7.136	5.895	5.111	4.565	4.161	3.848	3.598	3.393	2.739	2.383
	87	15.970	9.650	7.303	6.047	5.252	4.699	4.289	3.971	3.716	3.507	2.840	2.475
	127	16.233	9.856	7.484	6.213	5.408	4.847	4.431	4.108	3.850	3.637	2.958	2.584
	247	16.513	10.079	7.682	6.395	6.580	5.012	4.591	4.264	4.002	3.786	3.097	2.717
	∞	16.812	10.319	7.897	6.596	5.772	5.198	4.772	4.442	4.177	3.959	3.264	2.882

(续)

$p=7$													
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	18	11.961	7.063	5.258	4.304	3.708	3.298	2.999	2.770	2.589	2.442	1.989	1.753
	20	12.184	7.243	5.411	4.437	3.827	3.406	3.098	2.861	2.674	2.522	2.049	1.802
	24	12.513	7.516	5.647	4.647	4.016	3.580	3.258	3.011	2.814	2.653	2.151	1.887
	28	12.744	7.714	5.821	4.803	4.160	3.713	3.382	3.127	2.924	2.757	2.535	1.956
	32	12.915	7.863	5.954	4.952	4.272	3.817	3.481	3.220	3.012	2.842	2.304	2.015
	38	13.102	8.030	6.104	5.063	4.401	3.939	3.596	3.330	3.117	2.942	2.388	2.088
	48	13.308	8.216	6.275	5.222	4.551	4.081	3.732	3.461	3.243	3.063	2.492	2.180
	68	13.534	8.426	6.471	5.407	4.727	4.251	3.895	3.619	3.396	3.213	2.625	2.300
	88	13.657	8.541	6.579	5.511	4.827	4.348	3.990	3.711	3.486	3.301	2.706	2.376
	128	13.786	8.665	6.697	5.624	4.937	4.455	4.095	3.814	3.588	3.401	2.800	2.465
	248	13.923	8.797	6.824	5.747	5.058	4.573	4.211	3.929	3.702	3.514	2.910	2.573
	∞	14.067	8.938	6.961	5.882	5.191	4.705	4.343	4.060	3.833	3.645	3.041	2.705
.01	18	13.874	7.872	5.744	4.640	3.960	3.498	3.163	2.908	2.707	2.545	2.505	1.797
	20	14.310	8.164	5.971	4.829	4.124	3.642	3.292	3.026	2.816	2.646	2.124	1.855
	24	14.974	8.619	6.332	5.133	4.389	3.879	3.507	3.222	2.997	2.814	2.250	1.956
	28	15.456	8.957	6.605	5.367	4.595	4.065	3.676	3.379	3.143	2.951	2.355	2.042
	32	15.822	9.218	6.818	5.551	4.759	4.214	3.814	3.506	3.262	3.063	2.443	2.115
	38	16.230	9.514	7.063	5.765	4.952	4.390	3.977	3.659	3.406	3.199	2.551	2.206
	48	16.688	9.852	7.346	6.015	5.179	4.600	4.173	3.843	3.581	3.366	2.688	2.324
	68	17.206	10.243	7.679	6.313	5.452	4.855	4.413	4.072	3.799	3.575	2.866	2.481
	88	17.491	10.461	7.867	6.483	5.610	5.003	4.554	4.207	3.929	3.701	2.976	2.581
	128	17.796	10.697	8.073	6.670	5.785	5.169	4.713	4.360	4.078	3.846	3.106	2.700
	248	18.124	10.954	8.298	6.878	5.980	5.356	4.894	4.535	4.248	4.012	3.260	2.847
	∞	18.475	11.233	8.546	7.108	6.200	5.567	5.099	4.736	4.446	4.207	3.447	3.030

(续)

						p = 8							
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
	19	13.101	7.640	5.645	4.594	3.941	3.493	3.166	2.916	2.719	2.559	2.067	1.812
	21	13.346	7.834	5.808	7.737	4.067	3.607	3.270	3.012	2.808	2.643	2.130	1.863
	25	13.710	8.132	6.063	4.962	4.270	3.792	3.441	3.171	2.956	2.782	2.238	1.952
	29	13.970	8.350	6.253	5.131	4.425	3.935	3.574	3.295	3.074	2.893	2.326	2.025
	33	14.163	8.515	6.399	5.264	4.547	4.049	3.680	3.396	3.469	2.984	2.400	2.088
.05	39	14.377	8.701	6.566	5.416	4.688	4.181	3.806	3.515	3.283	3.092	2.490	2.165
	49	14.614	8.912	6.757	5.593	4.854	4.338	3.955	3.658	3.420	3.224	2.603	2.264
	69	14.877	9.151	6.977	5.800	5.050	4.526	4.136	3.832	3.589	3.388	2.747	2.395
	89	15.021	9.283	7.101	5.917	5.163	4.634	4.241	3.935	3.689	3.486	2.836	2.478
	129	15.173	9.426	7.235	6.046	5.287	4.755	4.358	4.050	3.802	3.597	2.940	2756
	249	15.335	9.579	7.381	6.187	5.424	4.889	4.491	1.480	3.931	3.725	3.063	2.695
	∞	15.507	9.745	7.541	6.342	5.577	5.040	4.640	4.329	4.078	3.872	3.210	2.843
	19	14.999	8.435	6.119	4.921	4.185	3.685	3.323	3.048	2.832	2.658	2.125	1.853
	21	15.463	8.743	6.357	5.119	4.355	3.836	3.458	3.171	2.944	2.762	2.201	1.913
	25	16.177	9.226	6.739	5.439	4.634	4.084	3.682	3.376	3.134	2.938	2.332	2.018
	29	16.700	9.589	7.030	5.687	4.853	4.280	3.861	3.541	3.287	3.081	2.442	2.107
	33	17.100	9.871	7.259	5.885	5.028	4.439	4.007	3.676	3.414	3.200	2.535	2.184
.01	39	17.549	10.194	7.524	6.115	5.234	4.627	4.181	3.839	3.566	3.344	2.649	2.280
	49	18.058	10.565	7.833	6.387	5.480	4.854	4.392	4.037	3.754	3.523	2.795	2.405
	69	18.640	10.998	8.199	6.713	5.778	5.131	4.653	4.284	3.990	3.749	2.986	2.573
	89	18.962	11.242	8.408	6.901	5.952	5.294	4.808	4.432	4.132	3.886	3.105	2.680
	129	19.310	11.508	8.638	7.109	6.146	5.478	4.983	4.601	4.295	4.044	3.246	2.810
	249	19.684	11.798	8.891	7.341	6.364	5.685	5.183	4.794	4.483	4.228	3.415	2.970
	∞	20.090	12.117	9.173	7.601	6.610	5.922	5.413	5.019	4.703	4.445	3.622	3.171

(续)

							$p = 10$						
α	n \ m	1	2	3	4	5	6	7	8	9	10	15	20
.05	21	15.322	8.761	6.395	5.158	4.392	3.869	3.489	3.199	2.970	2.785	2.217	1.925
	23	15.604	8.979	6.577	5.315	4.531	3.994	3.602	3.303	3.067	2.875	2.285	1.980
	27	16.033	9.320	6.684	5.566	4.756	4.199	3.790	3.477	3.229	3.028	2.403	2.075
	31	16.344	9.573	7.082	5.759	4.931	4.359	3.939	3.616	3.360	3.151	2.500	2.156
	35	16.580	9.769	7.252	5.912	5.071	4.488	4.060	3.730	3.467	3.253	2.581	2.225
	41	16.843	9.992	7.448	6.090	5.234	4.641	4.203	3.865	3.596	3.376	2.682	2.311
	51	17.140	10.247	7.676	6.299	5.429	4.824	4.376	4.030	3.754	3.527	2.810	2.423
	71	17.476	10.543	4.944	6.548	5.663	5.047	4.590	4.235	3.952	3.719	2.977	2.572
	91	17.662	10.709	8.097	6.691	5.800	5.178	4.716	4.358	4.070	3.834	3.081	2.668
	131	17.861	10.890	8.265	6.850	5.952	5.324	4.858	4.496	4.206	3.967	3.204	2.783
	251	18.076	11.087	8.450	7.027	6.122	5.490	5.021	4.656	4.363	4.122	3.350	2.925
	∞	18.307	11.303	8.654	7.224	6.315	5.679	5.207	4.840	4.546	4.304	3.529	3.102
.01	21	17.197	9.534	6.851	5.470	4.624	4.051	3.636	3.322	3.075	2.877	2.271	1.962
	23	17.707	9.867	7.107	5.682	4.806	4.211	3.779	3.452	3.194	2.987	2.351	2.025
	27	18.505	10.399	7.523	6.029	5.107	4.478	7.021	3.672	3.397	3.175	2.491	2.137
	31	19.101	10.805	7.846	6.302	5.346	4.693	4.216	3.851	3.564	3.330	2.608	2.233
	35	19.562	11.125	8.103	6.522	5.541	4.868	4.376	4.000	3.702	3.460	2.709	2.315
	41	20.088	11.495	8.408	6.782	5.772	5.078	4.570	4.180	3.871	3.619	2.835	2.420
	51	20.692	11.928	8.761	7.093	6.052	5.335	4.808	4.403	4.081	3.819	2.996	2.558
	71	21.394	12.441	9.190	7.473	6.397	5.654	5.107	4.686	4.350	4.076	3.211	2.745
	91	21.790	12.735	9.439	7.695	6.601	5.845	5.287	4.857	4.515	4.234	3.347	2.867
	131	22.221	13.059	9.716	7.944	6.832	6.062	5.494	5.055	4.705	4.418	3.510	3.016
	251	22.692	13.417	10.025	8.225	7.094	6.310	5.732	5.285	4.928	4.636	3.707	3.201
	∞	23.209	13.816	10.373	8.545	7.395	6.598	6.010	5.556	5.193	4.895	3.952	3.438

表 B.5 样本量相等时协方差阵相等的修正似然比检验的分位数

$\Pr\{-2\ln\lambda^* \geqslant x\} = 0.05$

$n_g \backslash q$	2	3	4	5	6	7	8	9	10
					$p=2$				
3	12.18	18.70	24.55	30.09	35.45	40.68	45.81	50.87	55.87
4	10.70	16.65	22.00	27.07	31.97	36.76	41.45	46.07	50.64
5	9.97	15.63	20.73	25.56	30.23	34.79	39.26	43.67	48.02
6	9.53	15.02	19.97	24.66	29.19	33.61	37.95	42.22	46.45
7	9.24	14.62	19.46	24.05	28.49	32.82	37.07	41.26	45.40
8	9.04	14.33	19.10	23.62	27.99	32.26	36.45	40.57	44.65
9	8.88	14.11	18.83	23.30	27.62	31.84	35.98	40.06	44.08
10	8.76	13.94	18.61	23.05	27.33	31.51	35.61	36.65	43.64
					$p=3$				
5	19.2	30.5	41.0	51.0	60.7	70.3	79.7	89.0	98.3
6	17.57	28.24	38.06	47.49	56.68	65.69	74.58	83.37	92.09
7	16.59	26.84	36.29	45.37	54.21	62.89	71.45	79.91	88.29
8	15.93	25.90	35.10	43.93	52.54	60.99	69.33	77.56	85.72
9	15.46	25.22	34.24	42.90	51.34	59.62	67.79	75.86	83.86
10	15.11	24.71	33.59	42.11	50.42	58.58	66.62	74.57	82.45
11	14.83	24.31	33.08	41.50	49.71	57.76	65.71	73.56	81.35
12	14.61	23.99	32.67	41.01	49.13	57.11	64.97	72.75	80.46
13	14.43	23.73	32.33	40.60	48.66	56.57	64.37	72.08	79.72
					$p=4$				
6	30.07	48.63	65.91	82.6	98.9	115.0	131.0	–	–
7	27.31	44.69	60.90	76.56	91.89	107.0	121.9	137.0	152.0
8	25.61	42.24	57.77	72.78	87.46	101.9	116.2	130.4	144.6
9	24.46	40.56	55.62	70.17	84.42	98.45	112.3	126.1	139.8
10	23.62	39.34	54.05	68.27	82.19	95.91	109.5	122.9	136.3
11	22.98	38.41	52.85	66.81	80.49	93.95	107.3	120.5	133.6
12	22.48	37.67	51.90	65.66	79.14	92.41	105.5	118.5	131.5
13	22.08	37.08	51.13	64.73	78.04	91.16	104.1	117.0	129.7
14	21.75	36.59	50.50	63.96	77.14	90.12	103.0	115.7	128.3
15	21.47	36.17	49.97	63.31	76.38	89.25	102.0	114.6	127.1

(续)

$n_g\backslash q$	2	3	4	5	6	7	$n_g\backslash q$	2	3	4	5
			$p=5$						$p=6$		
8	39.29	65.15	89.46	113.0	–	–	10	49.95	84.43	117.0	–
9	36.70	61.40	84.63	107.2	129.3	151.5					
10	34.92	58.79	81.25	103.1	124.5	145.7	11	47.43	80.69	112.2	142.9
							12	45.56	77.90	108.6	138.4
11	33.62	56.86	78.76	100.0	120.9	141.6	13	44.11	75.74	105.7	135.0
12	32.62	55.37	76.83	97.68	118.2	138.4	14	42.96	74.01	103.5	132.2
13	31.83	54.19	75.30	95.81	116.0	135.9	15	42.03	72.59	101.6	129.9
14	31.19	53.24	74.06	94.29	114.2	133.8					
15	30.66	52.44	73.02	93.02	112.7	132.1	16	41.25	71.41	100.1	128.0
							17	40.59	70.41	98.75	126.4
16	30.21	51.77	72.14	91.95	111.4	130.6	18	40.02	69.55	97.63	125.0
							19	39.53	68.80	96.64	123.8
							20	39.11	68.14	95.78	122.7

表 B.6 球形检验的分位数的校正因子

$n\backslash p$	3	4	5	6	7	8
	5% 显著性水平					
4	1.217					
5	1.074	1.322				
6	1.038	1.122	1.383			
7	1.023	1.066	1.155	1.420		
8	1.015	1.041	1.088	1.180	1.442	
9	1.011	1.029	1.057	1.098	1.199	1.455
10	1.008	1.021	1.040	1.071	1.121	1.214
12	1.005	1.013	1.023	1.039	1.060	1.093
14	1.004	1.008	1.015	1.024	1.037	1.054
16	1.003	1.006	1.011	1.017	1.025	1.035
18	1.002	1.005	1.008	1.012	1.018	1.025
20	1.002	1.004	1.006	1.010	1.014	1.019
24	1.001	1.002	1.004	1.006	1.009	1.012
28	1.001	1.002	1.003	1.004	1.006	1.008
34	1.000	1.001	1.002	1.003	1.004	1.005
42	1.000	1.001	1.001	1.002	1.002	1.003
50	1.000	1.000	1.001	1.001	1.002	1.002
100	1.000	1.000	1.000	1.000	1.000	1.000
χ^2	11.0705	16.9190	23.6848	31.4104	40.1133	49.8018
	1% 显著性水平					
$n\backslash p$	3	4	5	6	7	8
4	1.266					
5	1.091	1.396				
6	1.046	1.148	1.471			
7	1.028	1.079	1.186	1.511		
8	1.019	1.049	1.103	1.213	1.542	
9	1.013	1.034	1.067	1.123	1.234	1.556
10	1.010	1.025	1.047	1.081	1.138	1.250
12	1.006	1.015	1.027	1.044	1.068	1.104
14	1.004	1.010	1.018	1.028	1.041	1.060
16	1.003	1.007	1.012	1.019	1.028	1.039
18	1.002	1.005	1.009	1.014	1.020	1.028
20	1.002	1.004	1.007	1.011	1.015	1.021
24	1.001	1.003	1.005	1.007	1.010	1.013
28	1.001	1.002	1.003	1.005	1.007	1.009
34	1.001	1.001	1.002	1.003	1.004	1.006
42	1.000	1.001	1.001	1.002	1.003	1.003
50	1.000	1.001	1.001	1.001	1.002	1.002
100	1.000	1.000	1.000	1.000	1.000	1.001
χ^2	15.0863	21.6660	29.1412	37.5662	46.9629	57.3421

表 B.7[1] 修正似然比检验 $\Sigma = \Sigma_0$ 的分位数

$\Pr\{-2\ln\lambda_1^* \geqslant x\} = 0.05$

n	5%	1%	n	5%	1%	n	5%	1%	n	5%	1%
	$p=2$			$p=3$			$p=5$			$p=6$	
2	13.50	19.95	4	18.8	25.6	9	32.5	40.0	12	40.9	49.0
3	10.64	15.56	5	16.82	22.68	10	31.4	38.6	13	40.0	47.8
4	9.69	14.13							14	39.3	47.0
5	9.22	13.42	6	15.81	21.23	11	30.55	37.51	15	38.7	46.2
			7	15.19	20.36	12	29.92	36.72			
6	8.94	13.00	8	14.77	19.78	13	29.42	36.09	16	38.22	45.65
7	8.75	12.73	9	14.47	19.36	14	29.02	35.57	17	37.81	45.13
8	8.62	12.53	10	14.24	19.04	15	28.68	35.15	18	37.45	44.70
9	8.52	12.38							19	37.14	44.32
10	8.44	12.26	11	14.06	18.80	16	28.40	34.79	20	36.87	43.99
			12	13.92	18.61	17	28.15	34.49	21	36.63	43.69
	$p=4$		13	13.80	18.45	18	27.94	34.23			
7	25.8	30.8	14	13.70	18.31	19	27.76	34.00	22	36.41	43.43
8	24.06	29.33	15	13.62	18.20	20	27.60	33.79	24	36.05	42.99
9	23.00	28.36							26	35.75	42.63
10	22.28	27.66							28	35.49	42.32
									30	35.28	42.07
11	21.75	27.13									
12	21.35	26.71									
13	21.03	26.38									
14	20.77	26.10									
15	20.56	25.87									
	$p=7$			$p=8$			$p=9$			$p=10$	
18	48.6	56.9	24	58.4	67.1	28	70.1	79.6	34	(82.3)	(92.4)
19	48.2	56.3	26	57.7	66.3	30	69.4	78.8	36	81.7	91.8
20	47.7	55.8	28	57.09	65.68				38	81.2	91.2
21	47.34	55.36	30	56.61	65.12	32	68.8	78.17	40	80.7	90.7
22	47.00	54.96				34	68.34	77.60			
			32	56.20	64.64	36	(67.91)	(77.08)	45	79.83	89.63
24	46.43	54.28	34	55.84	64.23	38	(67.53)	(76.65)	50	79.13	88.83
26	45.97	53.73	36	55.54	63.87	40	67.21	76.29	55	78.57	88.20
28	45.58	53.27	38	55.26	63.55				60	78.13	87.68
30	45.25	52.88	40	55.03	63.28	45	66.54	75.51	65	77.75	87.26
32	44.97	52.55				50	66.02	74.92			
34	44.73	52.27				55	65.61	74.44	70	77.44	86.89
						60	65.28	74.06	75	77.18	86.59

① 括号中的元素已经被内插或外插到Korin表中，p =变量个数，N =观测个数，$n = N - 1$，$\lambda_1^* = n\ln|\Sigma_0| - np - n\ln|S| + n\mathrm{tr}(S\Sigma_0^{-1})$，其中$S$是样本协方差阵.

参考文献

在每个参考文献末的方括号内给出了在本书中用到该参考文献的章节号.

Abramowitz, Milton, and Irene Stegun (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards. U. S. Government Printing Office, Washington, D. C. [8.5]

Abruzzi, Adam (1950), *Experimental Procedures and Criteria for Estimating and Evaluating Industrial Productivity*, doctoral dissertation, Columbia University Library. [9.7, 9.P]

Adrian, Robert(1808), Research concerning the probabilities of the errors which happen in making observations, etc., *The Analyst or Mathematical Museum*, **1**, 93-109. [1.2]

Ahn, S. K., and G. C. Reinsel (1988), Nested reduced-rank autoregressive models for multiple time series, *Journal of American Statistical Association*, **83**, 849-856. [12.7]

Aitken, A. C. (1937), Studies in practical mathematics, II. The evaluation of the latent roots and latent vectors of a matrix, *Proceedings of the Royal Society of Edinburgh*, **57**, 269-305. [11.4]

Amemiya, Yasuo, and T. W. Anderson (1990), Asymptotic chi-square tests for a large class of factor analysis models, *Annals of Statistics*, **18**, 1453-1463. [14.6]

Anderson, R. L., and T. A. Bancroft (1952), *Statistical Theory in Research*, McGraw-Hill, New York. [8.P]

Anderson, T. W. (1946a), The non-central Wishart distribution and certain problems of multivariate statistics, *Annals of Mathematical Statistics*, **17**, 409-431. (Correction, **35** (1964), 923-924.) [14.4]

Anderson, T. W. (1946b), Analysis of multivariate variance, unpublished. [10.6]

Anderson, T. W. (1950), Estimation of the parameters of a single equation by the limited-information maximum-likelihood method, *Statistical Inference in Dynamic Economic Models* (Tjalling C. Koopmans, ed.) John Wiley & Sons, Inc., New York. [12.8]

Anderson, T. W. (1951a), Classification by multivariate analysis, *Pyschometrika*, **16**, 31-50. [6.5, 6.9]

Anderson, T. W. (1951b), Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics*, **22**, 327-351. (Correction, *Annals of Statistics*, **8** (1980), 1400.) [12.6, 12.7]

Anderson, T. W. (1951c), The asymptotic distribution of certain characteristic roots and vectors, *Proceedings of the Second Berkeley symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), University of California, Berkeley, 105-130. [12.7]

Anderson, T. W. (1955a), Some statistical problems in relating experimental data to predicting performance of a production process, *Journal of the American Statistical Association*, **50**, 163-177. [8.P]

Anderson, T. W. (1955b), The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proceedings of the American Mathematical Society*, **6**, 170-176. [8.10]

Anderson, T. W. (1957), Maximum likelihood estimates for a multivariate normal distribution when some obervations are missing, *Journal of the American Statistical Association*, **52**, 676-687. [4.P]

Anderson, T. W. (1963a), Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics*, **34**, 122-148. [11.3, 11.6, 11.7, 13.5]

Anderson, T. W. (1963b), A test for equality of means when covariance matrices are unequal, *Annals of Mathematical Statistics*, **34**, 671-672. [5.P]

Anderson, T. W. (1965a), Some optimum confidence bounds for roots of determinantal equations, *Annals of Mathematical Statistics*, **36**, 468-488. [11.6]

Anderson, T. W. (1965b), Some properties of confidence regions and tests of parameters ln multivariate distributions (with discussion). *Proceedings of the IBM Scientific Computing Symposium in Statistics, October 21-23, 1963*, IBM Data Processing Division, White Plains, New York, 15-28. [11.6]

Anderson, T. W. (1969), Statistical inference for covariance matrices with linear structure, *Multivariate Analysis II* (P. R. Krishnaiah, ed.), Academic, New York, 55-66. [3.P]

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York. [8.11]

Anderson, T. W. (1973a), Asymptotically efficient estimation of covariance matrices with linear structure, *Annals of Statistics*, **1**, 135-141. [14.3]

Anderson, T. W. (1973b), An asymptotic expansion of the distribution of the Studentized classification statistic *W*, *Annals of Statistics*, **1**, 964-972. [6.6]

Anderson, T. W. (1973c), Asymptotic evaluation of the probabilities of misclassification by linear discriminant functions, *Discriminant Analysis and Applications* (T. Cacoullos, ed.), Academic, New York, 17-35. [6.6]

Anderson, T. W. (1974), An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system, *Journal of the American Statistical Association*, **69**, 565-573. (Correction, **71** (1976), 1010.) [12.7]

Anderson, T. W. (1976), Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics (with discussion), *Journal of the Royal Statistical Society B*, **38**, 1-36. [12.7]

Anderson, T. W. (1977), Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences, *Econometrica*, **45**, 509-518. [12.7]

Anderson, T. W. (1984a), Estimating linear statistical relationships, *Annals of Statistics*, **12**, 1-45. [10.6, 12.6, 12.7, 14.1]

Anderson, T. W. (1984b), Asymptotic distribution of an estimator of linear functional relationships, unpublished. [12.7]

Anderson, T. W. (1987), Multivariate linear relations, *Proceedings of the Second International Tampere Conference in Statistics* (Tarmo Pukkila and Simo Puntanen, eds.), Tampere, Finland, 9-36. [12.7]

Anderson, T. W. (1989a), The asymptotic distribution of the likelihood ratio criterion for testing rank in multivariate components of variance, *Journal of Multivariate Analysis*, **30**, 72-79. [10.6, 12.7]

Anderson, T. W. (1989b), The asymptotic distribution of characteristic roots and vectors in multi-

variate components of variance, *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin* (Leon Jay Gleser, Michael D. Perlman, S. James Press, and Allan R. Sampson, eds.), Springer Verlag, New York, 177-196. [13.6]

Anderson, T. W. (1989c), Linear latent variable models and covariance structures, *Journal of Econometrics*, **41**, 91-119. [Correction **43**(1990), 395.] [12.8]

Anderson, T. W. (1993), Nonnormal multivariate distributions: Inference based on elliptically contoured distributions, *Multivariate Analysis: Future Directions* (C. R. Rao, ed.), North-Holland, Amsterdam, 1-25. [3.6]

Anderson, T. W. (1994), Inference in linear models, *Proceedings of the International Symposium on Multivariate Analysis and its Applications* (T. W. Anderson, K. T. Fang, and I. Olkin, eds.), Institute of Mathematical Statistics, 1-20. [12.7, 12.8, 14.3]

Anderson, T. W. (1999a), Asymptotic theory for canonical correlation analysis, *Journal of Multivariate Analysis*, **70**, 1-29. [13.7]

Anderson, T. W. (1999b), Asymptotic distribution of the reduced rank regression estimator under general conditions, *Annals of Statistics*, **27**, 1141-1154. [13.7]

Anderson, T. W. (2002), Specification and misspecification in reduced rank regression, Sankhyā, **64A**, 1-13. [12.7]

Anderson, T. W., and Yasuo Amemiya (1988a), The asymptotic normal distribution of estimators in factor analysis under general conditions, *Annals of Statistics*, **16**, 759-771. [14.6]

Anderson, T. W., and Yasuo Amemiya (1988b), Asymptotic distributions in factor analysis and linear structural relations, *Proceedings of the International Conference on Advances in Multivariate Statistical Analysis* (S. Das Gupta and J. K. Ghosh, eds.), Indian Statistical Institute, Calcutta, 1-22. [14.6]

Anderson, T. W., and R. R. Bahadur (1962), Classification into two multivariate normal distributions with different covariance matrices, *Annais of Mathematical Statistics*, **33**, 420-431. [6.10]

Anderson, T. W., and Kai-Tai Fang (1990a), On the theory of multivariate elliptically contoured distributions and their applications, *Statistical Inference in Elliptically Contoured and Related Distributions* (Kai-Tai Fang and T. W. Anderson, eds.), Allerton Press, Inc., New York, 1-23. [4.5]

Anderson, T. W., and Kai-Tai Fang (1990b), Inference in multivariate elliptically contoured distributions based on maximum likelihood, *Statistical Inference in Elliptically Contoured and Related Distributions* (Kai-Tai Fang and F. W. Anderson, eds.), Allerton Press, Inc., New York, 201-216. [3.6]

Anderson, T. W., and Kai-Tai Fang (1992), Theory and applications of elliptically contoured and related distributions, *The Development of Statistics: Recent Contributions from China* (X. R. Chen, K. T. Fang, and C. C. Yang, eds.), Longman Scientific and Technical, Harlow, Essex, 41-62. [10.11]

Anderson, T. W., Kai-Tai Fang, and Huang Hsu (1986), Maximum likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions, *Canadian Journal of Statistics*, **14**, 55-59. [Reprinted in *Statistical Inference in Elliptically Contoured and Related Distributions* (Kai-Tai Fang and T. W. Anderson, eds.), Allerton Press, Inc., New York, 1990, 217-223. [3.6]

Anderson, T. W., Somesh Das Gupta, and George P. H. Styan (1972), *A Bibliography of Multivariate Statistical Analysis*, Oliver & Boyd, Edinburgh. (Reprinted by Robert E. Krieger, Malabar, Florida, 1977.) [Preface]

Anderson, T. W., Naoto Kunitomo, and Takamitsu Sawa (1983a), Evaluation of the distribution function of the limited information maximum likelihood estimator, *Econometrica*, **50**, 1009-1027. [12.7]

Anderson, T. W., Naoto Kunitomo, and Takamitsu Sawa (1983b), Comparison of the densities of the TSLS and LIMLK estimators, *Global Econometrics, Essays in Honor of Lawrence R. Klein* (F. Gerard Adams and Bert Hickman, eds.), MIT, Cambridge, MA, 103-124. [12.7]

Anderson, T. W., and I. Olkin (1985), Maximum likelihood estimation of the parameters of a multivariate normal distribution, *Linear Algebra and Its Applications*, **70**, 147-171. [3.2]

Anderson, T. W., and Michael D. Perlman (1987), Consistency of invariant tests for the multivariate analysis of variance, *Proceedings of the Second International Tampere Conference in Statistics* (Tarmo Pukkila and Simo Puntanen, eds.) Tampere, Finland, 225-243. [8.10]

Anderson, T. W., and Michael D. Perlman (1993), Parameter consistency of invariant tests for MANOVA and related multivariate hypotheses. *Statistics and Probabifity: A Raghu Raj Bahadur Festschift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy, and B. L. S. Prakasa Rao, eds.), Wiley Eastern Limited, 37-62. [8.10]

Anderson, T. W., and Herman Rubin (1949), Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics*, **20**, 46-63. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, eds.), Colorado Associated University, 1970, 358-375.] [12.8]

Anderson, T. W., and Herman Rubin (1950), The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics*, **21**, 570-582. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, eds.), Colorado Associated University, 1970, 376-388.] [12.8]

Anderson, T. W., and Herman Rubin (1956), Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), Vol. V, University of California, Berkeley and Los Angeles, 111-150. [14.2, 14.3, 14.4, 14.6]

Anderson, T. W., and Takamitsu Sawa (1973), Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions, *Econometrica*, **41**, 683-714. [12.7]

Anderson, T. W., and Takamitsu Sawa (1977), Tables of the distribution of the maximum likelihood estimate of the slope coefficient and approximations, Technical Report No. 234, Economics Series, Institute for Mathematical Studies in the Social Sciences, Stanford University, April. [12.7]

Anderson, T. W., and Takamitsu Sawa (1979), Evaluation of the distribution function of the two-stage least squares estimate, *Econometrica*, **47**, 163-182. [12.7]

Anderson, T. W., and Takamitsu Sawa (1982), Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient, *Journal of the Royal Statistical Society B*, **44**, 52-62. [12.7]

Anderson, T. W., and George P. H. Styan (1982), Cochran's theorem, rank additivity and tripotent

matrices, *Statistics and Probability: Essays in Honor of C. R. Rao* (G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, eds.), North-Holland, Amsterdam, 1-23. [7.4]

Anderson, T. W., and Akimichi Takemura (1982), A new proof of admissibility of tests in multivariate analysis, *Journal of Multivariate Analysis*, **12**, 457-468. [8.10]

Anderson, Steen A., David Madigan, and Michael D. Perlman (2001), Alternative Markov properties for chain graphs, *Scandinavian Journal of Statistics*, **28**, 33-85. [15.2]

Barnard, M. M. (1935), The secular variations of skull characters in four series of Egyptian skulls, *Annals of Eugenics*, **6**, 352-371. [8.8]

Barndorff-Nielson, O. E. (1978), *Information and Exponential Families in Statistical Theory*, John Wiley & Sons, New York. [15.5]

Barnes, E. W. (1899), The theory of the gamma function, *Messenger of Mathematics*. **29**, 64-129. [8.5]

Bartlett, M. S. (1934), The vector representation of a sample, *Proceedings of the Cambridge Philosophical Society*, **30**, 327-340. [8.3]

Bartlett, M. S. (1937a), Properties of sufficiency and statistical tests, *Proceedings of the Royal Society of London A*, **160**, 268-282. [10.2]

Bartlett, M. S. (1937b), The statistical conception of mental factors, *British Journal of Psychology*, **28**, 97-104. [14.7]

Bartlett, M. S. (1938), Further aspects of the theory of multiple regression, *Proceedings of the Cambridge Philosophical Society*, **34**, 33-40. [14.7]

Bartlett, M. S. (1939), A note on tests of significance in multivariate analysis, *Proceedings of the Cambridge Philosophical Society*, **35**, 180-185. [7.2, 8.6]

Bartlett, M. S. (1947), Multivariate analysis, *Journal of the Royal Statistical Society, Supplement*, **9**, 176-197. [8.8]

Bartlett, M. S. (1950), Tests of significance in factor analysis, *British Journal of Psychology (Statistics Section)*, **3**, 77-85. [14.3]

Basmann, R. L. (1957), A generalized classical method of linear estimation of coefficients in a structural equation, *Econometrica*, **25**, 77-83. [12.8]

Basmann, R. L. (1961), A note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases, *Journal of the American Statistical Association*, **56**, 619-636. [12.8]

Basmann, R. L. (1963), A note on the exact finite sample frequency functions of generalized classical linear estimators in a leading three-equation case. *Journal of the American Statistical Association*, **58**, 161-171. [12.8]

Bennett, B. M. (1951), Note on a solution of the generalized Behrens-Fisher problem, *Annals of the Institute of Statistical Mathematics*, **2**, 87-90. [5.5]

Berger, J. O. (1975), Minimax estimation of location vectors for a wide class of densities, *Annals of Statistics*, **3**, 1318-1328. [3.5]

Berger, J. O. (1976), Admissibility results for generalized Bayes estimators of coordinates of a location vector, *Annals of Statistics*, **4**, 334-356. [3.5]

Berger, J. O. (1980a), A robust generalized Bayes estimator and confidence region for a multivariate normal mean, *Annals of Statistics*, **8**, 716-761. [3.5]

Berger, J. O. (1980b), *Statistical Decision Theory, Foundations, Concepts and Methods*, Springer-

Verlag, New York. [3.4, 6.2, 6.7]

Berndt, Ernst R., and Eugene Savin (1977), Conflict among criteria for testing hypotheses in the multivariate linear regression model, *Econometrica*, **45**, 1263-1277. [8.6]

Bhattacharya, P. K. (1966), Estimating the mean of a multivariate normal population with general quadratic loss function, *Annals of Mathematical Statistics*, **37**, 1819-1824. [3.5]

Birnbaum, Allan (1955), Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests, *Annals of Mathematical Statistics*, **26**, 21-36. [5.6, 8.10]

Björck, A., and G. Golub (1973), Numerical methods for computing angles between linear subspaces, *Mathematics of Computation*, 27, 579-594. [12.3]

Blackwell, David, and M. A. Girshick (1954), *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York. [6.2, 6.7]

Blalock, H. M., Jr (ed.) (1971), *Causal Models in the Social Sciences*, Aldine-Atheston, Chicago. [15.1]

Bonnesen, T., and W. Fenchel(1948), *Theorie der Konvexen Körper*, Chelsea, New York. [8.10]

Bose, R. C. (1936a), On the exact distribution and moment-coefficients of the D^2-statistic, *Sankyā*, **2**, 143-154. [11.3]

Bose, R. C. (1936b), A note on the distribution of differences in mean values of two samples drawn from two multivariate normally distributed populations, and the definition of the D^2-statistic, *Sankhyā*, **2**, 379-384. [3.3]

Bose, R. C., and S. N. Roy (1938), The distribution of the studentised D^2-statistic, *Sankhyā*, **4**, 19-38. [5.4]

Bowker, A. H. (1960), A representation of Hotclling's T^2 and Anderson's classification statistic W in terms of simple statistics, *Contributions to Probability and Statistics* (I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, eds.), Stanford University, Stanford, California, 142-149. [5.2]

Bowker, A. H., and R. Stigreaves (1961), An asymptotic expansion for the distribution function of the W-classification statistic, *Studies in Item Analysis and Prediction* (Herbert Solomon, ed.), Stanford University, Stanford, California, 293-310. [6.6]

Box, G. E. P. (1949), A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317-346. [8.5, 9.4, 10.4, 10.5]

Bravais, Auguste (1846), Analyse mathématique sur les probabilités des erreurs de situation d'un point, *Mémoires Présentés par Divers Savants à l'Académie Royale des Sciences de l'Institut de France*, **9**, 255-332. [1.2]

Brown, G. W. (1939), On the power of the L_1 test for equality of several variances, *Annals of Mathematical Statistics*, **10**, 119-128. [10.2]

Browne, M. W. (1974), Generalized least squares estimates in the analysis of covariance structures, *South African Statistical Journal*, **8**, 1-24. [14.3]

Chambers, John M. (1977), *Computational Methods for Data Analysis*, John Wiley & Sons, New York. [12.3]

Chan, Tony F., Gene H. Golub, and Randall J. LeVeque (1981), Algorithms for computing the sample variance: analysis and recommendations, unpublished. [3.2]

Chernoff, Herman (1956), Large sample theory: parametric case, *Annals of Mathematical Statistics*,

27, 1-22. [8.6]

Chernoff, Herman, and N. Divinsky (1953), The computation of maximum likelihood estimates of linear structural equations, *Studies in Econometric Method* (W. C. Hood and T. C. Koopmans, eds.), John Wiley & Sons, Inc., New York. [12.8]

Chu, S. Sylvia, and K. C. S. Pillai (1979), Power comparisons of two-sided tests of equality of two covariance matrices based on six criteria, *Annals of the Institute of Statistical Mathematics*, **31**, 185-205. [10.6]

Chung, Kai Lai (1974), *A Course in Probability Theory*, 2nd ed., Academic, New York. [2.2]

Clemm, D. S., P. R. Krishnaiah, and V. B. Waikar (1973), Tables for the extreme roots of the Wishart matrix, *Journal of Statistical Computation and Simulation*, **2**, 65-92. [10.8]

Clunies-Ross, C. W., and R. H. Riffenburgh (1960), Geometry and linear discrimination, *Biometrika*, **47**, 185-189. [6.10]

Cochran, W. G. (1934), The distribution of quadratic forms in a normal system. with applications to the analysis of covariance, *Proceedings of the Cambridge Philosophical Society*, **30**, 178-191. [7.4]

Constantine, A. G. (1963), Some non-central distribution problems in multivariate analysis, *Annals of Mathematical Statistics*, **34**, 1270-1285. [12.4]

Constantine, A. G. (1966), The distribution of Hotelling's generalisecd T_0^2, *Annals of Mathematical Statistics*, **37**, 215-225. [8.6]

Consul, P. C. (1966), On the exact distributions of the likelihood ratio criteria for testing linear hypotheses about regression coefficients, *Annals of Mathematical Statistics*, **37**, 1319-1330. [8.4]

Consul, P. C. (1967a), On the exact distributions of likelihood ratio criteria for testing independence of sets of variates under the null hypothesis, *Annals of Mathematical Statistics*, **38**, 1160-1169. [9.3]

Consul, P. C. (1967b), On the exact distributions of the criterion W for testing sphericity and in a p-variate normal distribution, *Annals of Mathematical Statistics*, **38**, 1170-1174. [10.7]

Courant, R., and D. Hilbert (1953), *Methods of Mathematical Physics*, Interscience, New York. [9.10]

Cox, D. R., and N. Wermuth (1996), *Multivariate Dependencies*, Chapman and Hall, London. [15.1]

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University, Princeton. [2.6, 3.2, 3.4, 6.5, 7.2]

Dahm, P. Fred, and Wayne A. Fuller (1981), Generalized least squares estimation of the functional multivariate linear errors in variables model, unpublished. [14.3]

Daly, J. F. (1940), On the unbiased character of likelihood-ratio tests for independence in normal systems, *Annals of Mathematical Statistics*, **11**, 1-32. [9.2]

Darlington, R. B., S. L. Weinberg, and H. J. Walberg (1973), Canonical variate analysis and related techniques, *Review of Educational Research*, **43**, 433-454. [12.2]

Das Gupta, Somesh (1965), Optimum classification rules for classification into two multivariate normal populations, *Annals of Mathematical Statistics*, **36**, 1174-1184. [6.6]

Das Gupta, S., T. W. Anderson, and G. S. Mudholkar (1964), Monotonicity of the power functions of some tests of the multivariate linear hypothesis, *Annals of Mathematical Statistics*, **35**,

200-205. [8.10]

David, F. N. (1937), A note on unbiased limits for the correlation coefficient, *Biometrika*, **29**, 157-160. [4.2]

David, F. N. (1938). *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples*, Cambridge University, Cambridge. [4.2]

Davis, A. W. (1968), A system of linear differential equations for the distribution of Hotelling's generalized T_0^2, *Annals of Mathematical Statistics*, **39**, 815-832. [8.6]

Davis, A. W. (1970a), Exact distributions of Hotelling's generalized T_0^2, *Biometrika*, **57**, 187-191. [Preface, 8.6]

Davis, A. W. (1970b), Further applications of a differential equation for Hotelling's generalized T_0^2, *Annals of the Institute of Statistical Mathematics*, **22**, 77-87. [Preface, 8.6]

Davis, A. W. (1971), Percentile approximations for a class of likelihood ratio criteria, *Biometrika*, **58**, 349-356. [10.8, 10.9]

Davis, A. W. (1972a), On the marginal distributions of the latent roots of the multivariate beta matrix, *Annals of Mathematical Statistics*, **43**, 1664-1670. [8.3]

Davis, A. W. (1972b), On the distributions of the latent roots and traces of certain random matrices, *Journal of Multivariate Analysis*, **2**, 189-200. [8.6]

Davis, A. W. (1980), Further tabulation of Hotelling's generalized T_0^2, *Communications in Statistics*, **B9**, 321-336. [Preface, 8.6]

Davis, A. W., and J. B. F. Field (1971), Tables of some multivariate test criteria. Technical Report No. 32, Division of Mathematical Statistics, C. S. I. R. O., Canberra, Australia. [10.8]

Davis, Harold T. (1933), *Tables of the Higher Mathematical Functions*, Vol. I, Principia Press, Bloomington, Indiana. [8.5]

Davis, Harold T. (1935), *Tables of the Higher Mathematical Functions*, Vol. II. Principia Press, Bloomington, Indiana. [8.5]

Deemer, Walter L., and Ingram Olkin (1951), The Jacobians of certain matrix transformations useful in multivariate analysis. Based on lectures of P. L. Hsu at the University of North Carolina, 1947, *Biometrika*, **38**, 345-367. [13.2]

De Groot, Morris H. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York. [3.4, 6.2, 6.7]

Dempster, A. P. (1972), Covariance selection, *Biometrics*, **28**, 157-175. [15.5]

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statical Society B*, **39**, 1-38. [14.3]

Diaconis, Persi, and Bradley Efron (1983), Computer- intensive methods in statistics. *Scientific American*, **248**, 116-130. [4.3]

Eaton, M. L., and M. D. Perlman (1974), A monotonicity property of the power functions of some invariant tests for MANOVA, *Annals of Statistic*, **2**, 1022-1028. [8.10]

Edwards, D. (1995), *Introduction to Graphical Modelling*, Springer-Verlag, New York. [15.1]

Efron, Bradley (1982), *The Jackknife, the Bootstrap, and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia. [4.2]

Efron, Bradley, and Carl Morris (1977), Stein's paradox in statistics, *Scientific American*, **236**, 119-127. [3.5]

Elfving, G. (1947), A simple method of deducing certain distributions connected with multivariate

sampling, *Skandinavisk Aktuarietidskrifi*, **30**, 56-74. [7.2]

Fang, Kai-Tai, Samuel Kotz, and Kai-Wang Ng (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, New York. [2.7]

Fang, Kai-Tai, and Yaó-Ting Zhang (1990), *Generalized Multivariate Analysis*, Springer-Verlag, New York. [2.7, 3.6, 10.11, 7.9]

Ferguson, Thomas Shelburne (1967), *Mathematical Statbstics: A Decision Theoretic Approach*, Academic, New York. [3.4, 6.2, 6.7]

Fisher, R. A. (1915), Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, *Biometrika*, **10**, 507-521. [4.2, 7.2]

Fisher, R. A. (1921), On the "probable error" of a coefficient of correlation deduced from a small sample, *Metron*, **1**, Part 4, 3-32. [4.2]

Fisher, R. A. (1924), The distribution of the partial correlation coefficient, *Metron*, **3**, 329-332. [4.3]

Fisher, R. A. (1928), The general sampling distribution of the multiple correlation coefficient, *Proceedings of the Royal Society of London, A*, **121**, 654-673. [4.4]

Fisher, R. A. (1936), The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, **7**, 179-188. [5.3, 6.5, 11.5]

Fisher, R. A. (1939), The sampling distribution of some statistics obtained from non-linear equations, *Annals of Eugenics*, **9**, 238-249. [13.2]

Fisher, R. A. (1947a), *The Design of Experiments* (4th ed.), Oliver and Boyd, Edinburgh. [8.9]

Fisher, R. A. (1947b), The analysis of covariance method for the relation between a part and the whole, *Biometrics*, **3**, 65-68. [3.P]

Fisher, R. A., and F. Yates (1942), *Statistical Tables for Biological, Agricultural and Medical Research* (2nd ed.), Oliver and Boyd, Edinburgh. [4.2]

Fog, David (1948), The geometrical method in the theory of sampling, *Biometrika*, **35**, 46-54. [7.2]

Foster, F. G. (1957), Upper percentage points of the generalized beta distribution, II, *Biometrika*, **44**, 441-453. [8.6]

Foster, F. G. (1958), Upper percentage points of the generalized beta distribution, III, *Biometrika*, **45**, 492-503. [8.6]

Foster, F. G., and D. H. Rees (1957), Upper percentage points of the generalized beta distribution, I, *Biometrika*, **44**, 237-247. [8.6]

Frets, G. P. (1921), Heredity of head form in man, *Genetica*, **3**, 193-384. [3.P]

Frisch, R. (1929), Correlation and scatter in statistical variables, *Nordic Statistical Journal*, **8**, 36-102. [7.5]

Frydenberg, M. (1990), The chain graph Markov property, *Scandinavian Journal of Statistics*, **17**, 333-353. [15.2]

Fujikoshi, Y. (1973), Monotonicity of the power functions of some tests In general MANOVA models, *Annals of Statistics*, **1**, 388-391. [8.6]

Fuiikoshi, Y. (1974), The likelihood ratio tests for the dimensionality of regression coefficients, *Journal of Multivariate Analysis*, **4**, 327-340. [12.4]

Fuiikoshi, Y., and M. Kanazawa (1976), The ML classification statistic in covariate discriminant analysis and its asymptotic expansions, *Essays in Probability and Statics*, 305-320. [6.6]

Fuller, Wayne A., Sastry, G. Pantula, and Yasuo Amemiya (1982), The covariance matrix of esti-

mators for the factor model, unpublished. [14.4]

Gabriel, K. R. (1969), Simultaneous test procedures–some theory of multiple comparisons, *Annals of Mathematical Statistics*, **40**, 224-250. [8.7]

Gaijar, A. V. (1967), Limiting distributions of certain transformations of multiple correlation coefficients, *Metron*, **26**, 189-193. [4.4]

Galton, Francis(1889), *Natural Inheritance*, MacMillan, London. [1.2, 2.5]

Gauss, K. F. (1823), *Theory of the Combination of Observations*, Göttingen. [1.2]

Giri, N. (1977), *Multivariate Statistical Inference*, Academic, New York. [7.2, 7.7]

Giri, N., and J. Kiefer (1964), Local and asymptotic minimax properties of multivariate tests. *Annals of Mathematical Statistics*, **35**, 21-35. [5.6]

Giri, N., J. Kiefer, and C. Stein (1963), Minimax character of Hotelling's T^2 test in the simplest case, *Annals of Mathematical Statistics*, **34**, 1524-1535. [5.6]

Girshick, M. A. (1939), On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, **10**, 203-224. [13.2]

Gleser, Leon Jay (1981), Estimation in a multivariate"errors in variables " regression model: large sample results, *Annals of Statistics*, **9**, 24-44. [12.7]

Glynn, W. J., and R. J. Muirhead (1978), Inference in canonical correlation analysis, *Journal of Multivariate Analysis*, **8**, 468-478. [12.4]

Golub, Gene H., and Franklin T. Luk (1976), Singular value decomposition: applications and computations, unpublished. [12.3]

Golub, Gene H., and Charles F. Van Loan (1989), *Matrix Computations* (2nd ed.), Johns Hopkins University Press, Baltimore. [11.4, 12.3, A.5]

Grubbs, F. E. (1954), Tables of 1% and 5% probability levels of Hotelling's generalized T^2 statistics, Technical Note No. 926, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland. [8.6]

Gupta, Shanti S. (1963), Bibliography on the multivariate normal integrals and related topics, *Annals of Mathematical Statistics*, **34**, 829-838. [2.3]

Gurland, John (1968), A relatively simple form of the distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society B*, **30**, 276-283. [4.4]

Gurland, J., and R. Milton (1970), Further consideration of the distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society B*, **32**, 381-394. [4.4]

Haavelmo, T. (1944), The probability approach in econometrics, *Econometrica*, **12**, Supplement, 1-118. [12.7]

Haff, L. R. (1980), Empirical Bayes estimation of the multivariate normal covariance matrix, *Annals of Statistics*, **8**, 586-597. [7.8]

Halmos, P. R. (1950), *Measure Theory*, D. van Nostrand, New York. [4.5, 13.3]

Harris, Bernard, and Andrew P. Soms (1980), The use of the tetrachoric series for evaluating multivariate normal probabilities, *Journal of Multivariate Analysis*, **10**, 252-267. [2.3]

Hayakawa, Takesi (1967), On the distribution of the maximum latent root of a positive definite symmetric random matrix, *Annals of the Institute of Statistical Mathematics*, **19**, 1-17. [8.6]

Heck, D. L. (1960), Charts of some upper percentage points of the distribution of the largest characteristic root, *Annals of Mathematical Statistics*, **31**, 625-642. [8.6]

Hickman, W. Braddock (1953), *The Volume of Corporate Bond Financing Since 1900*, Princeton

University, Princeton, 82-90. [10.7]

Hoel, Paul G. (1937), A significance test for component analysis, *Annals of Mathematical Statistics*, **8**, 149-158. [7.5]

Hooker, R. H. (1907), The correlation of the weather and crops, *Journal of the Royal Statistical Society*, **70**, 1-42. [4.2]

Hotelling, Harold (1931), The generalization of Student's ratio, *Annals of Mathematical Statistics*, **2**, 360-378. [5.1, 5.P]

Hotelling, Harold (1933), Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, **24**, 417-441, 498-520. [11.1, 14.3]

Hotelling, Harold (1935), The most predictable criterion, *Journal of Educational Psychology*, **26**, 139-142. [12.2]

Hotelling, Harold (1936), Relations between two sets of variates, *Biometrika*, **28**, 321-377. [12.1]

Hotelling, Harold(1947), Multivariate quality control, illustrated by the air testing of sample bombsights, *Techniques of Statistical Analysis* (C. Eisenhart, M. Hastay, and W. A. Wallis, eds.), McGraw-Hill, New York, 111-184. [8.6]

Hotelling, Harold (1951), A generalized T test and measure of multivariate dispersion, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), University of California, Los Angeles and Berkeley, 23-41. [8.6, 10.7]

Hotelling, Harold(1953), New light on the correlation coefficient and its transforms (with discussion), *Journal of the Royal Statistical Society B*, **15**, 193-232. [4.2]

Howe, W. G. (1955), *Some Contributions to Factor Analysis*, U.S. Atomic Energy Commission Report, Oak Ridge National Laboratory, Oak Ridge, Tennessee. [14.2, 14.6]

Hsu, P. L. (1938), Notes on Hotelling's generalized T, *Annals of Mathematical Statistics*, **9**, 231-243. [5.4]

Hsu, P. L. (1939a), On the distribution of the roots of certain determinantal equations, *Annals of Eugenics*, **9**, 250-258. [13.2]

Hsu, P. L. (1939b), A new proof of the joint product moment distribution, *Proceedings of the Cambridge Philosophical Society*, **35**, 336-338. [7.2]

Hsu, P. L. (1945), On the power functions for the E^2-test and the T^2-test, *Annals of Mathematical Statistics*, **16**, 278-286. [5.6]

Hudson, M. (1974), Empirical Bayes estimation, Technical Report NO. 58, NSF contract GP 30711X-2, Department of Statistics, Stanford University. [3.5]

Immer, F. R., H. D. Hayes, and LeRoy Powers (1934), Statistical determination of barley varietal adaptation, *Journal of the American Society of Agronomy*, **26**, 403-407. [8.9]

Ingham, A. E. (1933), An integral which occurs in statistics, *Proceedings of the Cambridge Philosophical Society*, **29**, 271-276. [7.2]

Ito, K. (1956), Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic, *Annals of Mathematical Statistics*, **27**, 1091-1105. [8.6]

Ito, K. (1960), Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic, II, *Annals of Mathematical Statistics*, **31**, 1148-1153. [8.6]

Izenman, A. J. (1975), Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis*, **5**, 248-264. [12.7]

Izenman, Alan Julian (1980), Assessing dimensionality in multivariate regression, *Analysis of Vari-*

ance, Handbook of Statistics, Vol. 1(P. R. Krishnaiah, ed.), North-Holland, Amsterdam, 571-591. [3.P]

James, A. T. (1954), Normal multivariate analysis and the orthogonal group, *Annals of Mathematical Statistics*, **25**, 40-75. [7.2]

James, A. T. (1964), Distributions of matrix variates and latent roots derived from normal samples, *Annals of Mathematical Statistics*, **35**, 475-501. [8.6]

James, W., and C. Stein (1961), Estimation with quadratic loss, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), Vol. I, 361-379. University of California, Berkeley. [3.5, 7.8]

Japanese Standards Association (1972). *Statistical Tables and Formulas with Computer Applications.* [Preface]

Jennrich, Robert I., and Dorothy T. Thayer (1973), A note on Lawley's formulas for standard errors in maximum likelihood factor analysis, *Psychometrika*, **38**, 571-580. [14.3]

Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press. [12.7]

Jolicoeur, Pierre, and J. E. Mosimann (1960), Size and shape variation in the painted turtle, a principal component analysis, *Growth*, **24**, 339-354. Also in *Benchmark Papers in Systematic and Evolutionary Biology* (E. H. Bryant and W. R. Atchley, eds.), **2**(1975), 86-101. [11.P]

Jöreskog, K. G. (1969), A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika*, **34**, 183-202. [14.2]

Jöreskog, K. G., and Arthur S. Goldberger (1972), Factor analysis by generalized least squares, *Psychometrika*, **37**, 243-260. [14.3]

Kaiser, Henry F. (1958), The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, **23**, 187-200. [14.5]

Kanazawa, M. (1979), The asymptotic cut-off point and comparison of error probabilities in covariate discriminant analysis, *Journal of the Japan Statistical Society*, **9**, 7-17. [6.6]

Kelley, T. L. (1928), *Crossroads in the Mind of Man*, Stanford University, Stanford. [4.P, 9.P]

Kendall, M. G., and Alan Stuart(1973), *The Advanced Theory of Statistics* (3rd ed.), Vol. 2, Charles Griffin, London. [12.6]

Kennedy, William J., Jr., and James E. Gentle (1980), *Statistical Computing*, Marcel Dekker, New York. [12.3]

Khatri, C. G. (1963), Joint estimation of the parameters of multivariate normal populations, *Journal of Indian Statistical Association*, **1**, 125-133. [7.2]

Khatri, C. G. (1966), A note on a large sample distribution of a transformed multiple correlation coefficient, *Annals of the Institute of Statistical Mathematics*, **18**, 375-380. [4.4]

Khatri, C. G. (1972), On the exact finite series distribution of the smallest or the largest root of matrices in three situations, *Journal of Multivariate Analysis*, **2**, 201-207. [8.6]

Khatri, C. G., and K. C. Sreedharan Pillai (1966), On the moments of the trace of a matrix and approximations to its non-central distribution. *Annals of Mathematical Statistics*, **37**, 1312-1318. [8.6]

Khatri, C. G., and K. C. S. Pillai (1968), On the noncentral distributions of two test criteria in multivariate analysis of variance, *Annals of Mathematical Statistics*, **39**, 215-226. [8.6]

Khatri, C. G., and K. V. Ramachandran (1958), Certain multivariate distribution problems, I

(Wishart's distribution), *Journal of the Maharaja Sayajairo, University of Baroda*, **7**, 79-82. [7.2]

Kiefer, J. (1957), Invariance, minimax sequential estimation, and continuous time processes, *Annals of Mathematical Statistics*, **28**, 573-601. [7.8]

Kiefer, J. (1966), Multivariate optimality results, *Multivariate Analysis* (Parachuyi R. Krishnaiah, ed.), Academic, New York, 255-274. [7.8]

Kiefer, J., and R. Schwartz (1965), Admissible Bayes character of T^2-, R^2-, and other fully invariant tests for classical multivariate normal problems, *Annals of Mathematical Statistics*, **36**, 747-770. [5.P, 9.9, 10.10]

Klotz, Jerome, and Joseph Putter (1969), Maximum likelihood estimation of the multivariate covariance components for the balanced one-way layout. *Annals of Mathematical Statistics*, **40**, 1100-1105. [10.6]

Kolmogorov, A. (1950), *Foundations of the Theory of Probability*, Chelsea, New York. [2.2]

Konishi, Sadanori (1978a), An approximation to the distribution of the sample correlation coefficient, *Biometrika*, **65**, 654-656. [4.2]

Konishi, Sadanori (1978b), Asymptotic expansions for the distributions of statistics based on a correlation matrix, *Canadian Journal of Statistics*, **6**, 49-56. [4.2]

Konishi, Sadanori (1979), Asymptotic expansions for the distributions of functions of a correlation matrix, *Journal of Multivariate Analysis*, **9**, 259-266. [4.2]

Koopmans, T. C., and Olav Reiersøl (1950), The identification of structural characteristics, *Annals of Mathematical Statistics*, **21**, 165-181. [14.2]

Korin, B. P. (1968), On the distribution of a statistic used for testing a covariance matrix, *Biometrika*, **55**, 171-178. [10.8]

Korin, B. P. (1969), On testing the equality of k covariance matrices, *Biometrika*, **56**, 216-218. [10.5]

Kramer, K. H. (1963), Tables for constructing confidence limits on the multiple correlation coefficient, *Journal of the American Statistical Association*, **58**, 1082-1085. [4.4]

Krishnaiah, P. R. (1978), Some recent developments on real multivariate distributions, *Development in Statistics* (P. R. Krishnaiah, ed.), Vol. 1, Academic, New York, 135-169. [8.6]

Krishnaiah, P. R. (1980), Computations of some multivariate distributions, *Analysis of Variance, Handbook of Statistics*, Vol. 1 (P. R. Krishnaiah, ed.), North-Holland, Amsterdam, 745-971.

Krishnaiah, P. R., and T. C. Chang (1972), On the exact distributions of the traces of $S_1(S_1+S_2)^{-1}$ and $S_1S_2^{-1}$, *Sankhyā*, *A*, **34**, 153-160. [8.6]

Krishnaiah, P. R., and F. J. Schuurmann (1974), On the evaluation of some distributions that arise in simultaneous tests for the equality of the latent roots of the covariance matrix, *Journal of Multivariate Analysis*, **4**, 265-282. [10.7]

Kshirsagar, A. M. (1959), Bartlett decomposition and Wishart distribution, *Annals of Mathematical Statistics*, **30**, 239-241. [7.2]

Kudo, H. (1955), On minimax invariant estimates of the transformation parameter, *Natural Science Report*, **6**, 31-73, Ochanomizu University, Tokyo, Japan. [7.8]

Kunitomo, Naoto (1980), Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association*, **75**, 693-700. [12.7]

Lachenbruch, P. A., and M. R. Mickey (1968), Estimation of error rates in discriminant analysis, *Technometrics*, **10**, 1-11. [6.6]

Laplace, P. S. (1811), Mémoire sur les intégrales définies et leur application aux probabilités, *Mémoires de l'Institut lmpérial de France,Année 1810*, 279-347. [1.2]

Lauritzen, Steffen L. (1996), *Graphical Models*, Clarendon Press, Oxford. [15.5]

Lauritzen, Steffen L., and N. Wermuth (1989), Graphical models for associations between variables some of which are qualitative and some quantitative, *Annals of Statistics*, **17**, 31-57. [15.2]

Läuter, Jürgen, Ekkehard Glimm, and Siegfried Kropf (1996a), New multivariate tests for data with an inherent structure, *Biometrics Journal*, **38**, 5-23. [Correction: **40**, (1998), 1015.][5.7]

Läuter, Jürgen, Ekkehard Glimm, and Siegfried Kropf (1996b), Multivariate tests based on left-spherically distributed linear scores, *Annals of Statistics*, **26**, 1972-1988. [5.7]

Läuter, Jürgen, Ekkehard Glimm, and Siegfried Kropf (1996c), Exact stable multivariate tests for applications in clinical research. *ASA Proceedings of the Biopharma-ceutical Section*, 46-55. [5.7]

Lawley, D. N. (1938), A generalization of Fisher's z test, *Biometrika*, 30, 180-187. [8.6]

Lawley, D. N. (1940), The estimation of factor loadings by the method of maximum likelihood, *Proceedings of the Royal Society of Edinburgh, Sec. A*, **60**, 64-82. [14.3]

Lawley, D. N. (1941), Further investigations in factor estimation, *Proceedings of the Royal Society of Edinburgh, Sec. A*, **61**, 176-185. [14.4]

Lawley, D. N. (1953), A modified method of estimation in factor analysis and some large sample results, *Uppsala Symposium on Psychological Factor Analysis, 17-19 March 1953*, Uppsala, Almqvist and Wiksell, 35-42. [14.3]

Lawley, D. N. (1958), Estimation in factor analysis under various initial assumptions. *British Journal of Statistical Psychology*, **11**, 1-12. [14.2, 14.6]

Lawley, D. N. (1959), Tests of significance in canonical analysis, *Biometrika*, **46**, 59-66. [12.4]

Lawley, D. N. (1967), Some new results in maximum likelihood factor analysis. *Proceedings of the Royal Society of Edinburgh, Sec. A*, **87**, 256-264. [14.3]

Lawley, D. N., and A. E. Maxwell (1971), *Factor Analysis as a Statistical Method* (2nd ed.), American Elsevier, New York. [14.3, 14.5]

Lee, Y. S. (1971a), Asymptotic formulae for the distribution of a multivariate test statistic: power comparisons of certain multivariate tests, *Biometrika*, **58**, 647-651. [8.6]

Lee, Y. S. (1971b), Some results on the sampling distribution of the multiple correlation coefficient, *Journal of the Royal Statistical Society B*, **33**, 117-130. [4.4]

Lee, Y. S. (1972), Tables of upper percentage points of the multiple correlation coefficient, *Biometrika*, **59**, 175-189. [4.4]

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, John Wiley & Sons, New York. [4.2, 5.6]

Lehmer, Emma(1944), Inverse tables of probabilities of errors of the second kind, *Annals of Mathematical Statistics*, **15**, 388-398. [5.4]

Loève, M. (1977), *Probability Theory I* (4th ed.), Springer-Verlag, New York. [2. 2]

Loève, M. (1978), *Probability Theory II* (4th ed.), Springer-Verlag, New York. [2.2]

Madow, W. G. (1938), Contributions to the theory of multivariate statistical analysis, *Transactions of the American Mathematical Society*, **44**, 454-495. [7.2]

Magnus, Jan R. (1988), *Linear Structures*, Charles Griffin and Co., London. [A.4]

Magnus, J. R., and H. Neudecker (1979), The commutation matrix: some properties and applications, *The Annals of Statistics*, **7**, 381-394. [3.6]

Mahalanobis, P. C. (1930), On tests and measures of group divergence, *Journal and Proceedings of the Asiatic Society of Bengal*, **26**, 541-588. [3.3]

Mahalanobis, P. C., R. C. Bose, and S. N. Roy (1937), Normalisation of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions, *Sankhyā*, **3**, 1-40. [7.2]

Mallows, C. L. (1961), Latent vectors of random symmetric matrices, *Biometrika*, **48**, 133-149. [11.6]

Mardia, K. V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519-530. [3.6]

Mariano, Roberto S., and Takamitsu Sawa (1972), The exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables, *Journal of the American Statistical Association*, **67**, 159-163. [12.7]

Maronna, R. A. (1976), Robust M-estimators of multivariate location and scatter, *Annals of Statistics*, **4**, 51-67. [3.6]

Marshall, A. W., and I. Olkin (1979), *Inequalities: Theory of Majorization and Its Applications*, Academic, New York. [8.10]

Mathai, A. M. (1971), On the distribution of the likelihood ratio criterion for testing linear hypotheses on regression coefficients, *Annals of the Institute of Statistical Mathematics*, **23**, 181-197. [8.4]

Mathai, A. M., and R. S. Katiyar (1979), Exact percentage points for testing independence, *Biometrika*, **66**, 353-356. [9.3]

Mathai, A. M., and P. N. Rathie (1980), The exact non-null distribution for testing equality of covariance matrices, *Sankhyā A*, **42**, 78-87. [10.4]

Mathai, A. M., and R. K. Saxena (1973), *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Lecture Notes No. 348, Springer-Verlag, New York. [9.3]

Mauchly, J. W. (1940), Significance test for sphericity of a normal n-variate distribution, *Annals of Mathematical Statistics*, **11**, 204-209. [10.7]

Mauldon, J. G. (1955), Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory, *Journal of the Royal Statistical Society B*, **17**, 79-85. [7.2]

McDonald, Roderick P. (2002), What can we learn from path equations: identifiability, constraints, equivalence, *Psychometrika*, **67**, 225-249. [15.1]

McLachlan, G. J. (1973), An asymptotic expansion of the expectation of the estimated error rate in discriminant analysis, *Australian Journal of Statistics*, **15**, 210-214. [6.6]

McLachlan, G. J. (1974a), An asymptotic unbiased technique for estimating the error rates in discriminant analysis, *Biometrics*, **30**, 239-249. [6.6]

McLachlan, G. J. (1974b), Estimation of the errors of misclassification on the criterion of asymptotic mean square error, *Technometrics*, **16**, 255-260. [6.6]

McLachlan, G. J. (1974c), The asymptotic distributions of the conditional error rate and risk in discriminant analysis, *Biometrika*, **61**, 131-135. [6.6]

McLachlan, G. J. (1977), Constrained sample discrimination with the studentized classification statistic W *Communications in Statistics-Theory and Methods*, **A6**, 575-583. [6.6]

Memon, A. Z., and M. Okamoto (1971), Asymptotic expansion of the distribution of The Z statistic

in discriminant analysis, *Journal of Multivariate Analysis*, **1**, 294-307. [6.6]

Mijares, T. A. (1964), *Percentage Points of the Sum $V_1^{(s)}$ of s Roots*(s=1−50), The Statistical Center, University of the Philippines, Manila. [8.6]

Mikhail, N. N. (1965), A comparison of tests of the Wilks-Lawley hypothesis in multivariate analysis, *Biometrika*, **52**, 149-156. [8.6]

Mood, A. M. (1951), On the distribution of the characteristic roots of normal second-moment matrices, *Annals of Mathematical Statistics*, **22**, 266-273. [13.2]

Morris, Blair, and Ingram Olkin (1964), Some estimation and testing problems for factor analysis models, unpublished. [10.6]

Mudholkar, G. S. (1966), On confidence bounds associated with multivariate analysis of variance and non-independence between two sets of variates, *Annals of Mathematical Statistics*, **37**, 1736-1746. [8.7]

Mudholkar, Govind S., and Madhusudan C. Trivedi (1980), A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance, *Biometrika*, **67**, 485-488. [8.5]

Mudholkar, Govind S., and Madhusudan C. Trivedi (1981), A normal approximation for the multivariate likelihood ratio statistics, *Statistical Distributions in Scientific Work* (C. Taillie et al., eds.), Vol. 5, 219-230, D. Reidel Publishing. [8.5]

Muirhead, R. J. (1970), Asymptotic distributions of some multivariate tests, *Annals of Mathematical Statistics*, **41**, 1002-1010. [8.6]

Muirhead, R. J. (1980), The effects of elliptical distributions on some standard procedures involving correlation coefficients: a review, *Multivariate Statistical Analysis* (R. P. Gupta, ed.), 143-159. [4.5]

Muirhead, Robb J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York. [2.7, 3.6, 7.7]

Muirhead, R. J., and C. M. Waternaux (1980), Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika*, **67**, 31-43. [4.5]

Nagao, Hisao (1973a), On some test criteria for covariance matrix, *Annals of Statistics*, **1**, 700-709. [9.5, 10.2, 10.7, 10.8]

Nagao, Hisao (1973b), Asymptotic expansions of the distributions of Bartlett's test and sphericity test under the local alternatives, *Annals of the Institute of Statistical Mathematics*, **25**, 407-422. [10.5, 10.6]

Nagao, Hisao (1973c), Nonnull distributions of two test criteria for independence under local alternatives, *Journal of Multivariate Analysis*, **3**, 435-444. [9.4]

Nagarsenker, B. N., and K. C. S. Pillai (1972), *The Distribution of the. Sphericity Test Criterion*, ARL 72-0154, Aerospace Research Laboratories. [Preface]

Nagarsenker, B. N., and K. C. S. Pillai (1973a), The distribution of the sphericity test criterion, *Journal of Multivariate Analysis*, **3**, 226-235. [10.7]

Nagarsenker, B. N., and K. C. S. Pillai (1973b), Distribution of the likelihood ratio criterion for testing a hypothesis specifying a covariance matrix, *Biometrika*, **60**, 359-364. [10.8]

Nagarsenker, B. N., and K. C. S. Pillai (1974), Distribution of the likelihood ratio criterion for testing $\Sigma = \Sigma_0$, $\mu = \mu_0$, *Journal of Multivariate Analysis*, **4**, 114-122. [10.9]

Nanda, D. N. (1948), Distribution of a root of a determinantal equation, *Annals of Mathematical Statistics*, **19**, 47-57. [8.6]

Nanda, D. N. (1950), Distribution of the sum of roots of a determinantai equation under a certain condition, *Annals of Mathematical Statistics*, **21**, 432-439. [8.6]

Nanda, D. N. (1951), Probability distribution tables of the larger root of a determinantal equation with two roots, *Journal of the Indian Society of Agricultural Statistics*, **3**, 175-177. [8.6]

Narain, R. D. (1948), A new approach to sampling distributions of the multivariate normal theory, I, *Journal of the Indian Society of Agricultural Statistics*, **1**, 59-69. [7.2]

Narain, R. D. (1950), On the completely unbiased character of tests of independence in multivariate normal systems, *Annals of Mathematical Statistics*, **21**, 293-298. [9.2]

National Bureau of Standards, United States (1959), *Tables of the Bivariate Normal Distribution Function and Related Functions*, U. S. Government Printing Office, Washington, D. C. [2.3]

Neveu, Jacques (1965), *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco. [2.2]

Ogawa, J. (1953), On the sampling distributions of classical statistics in multivariate analysis, *Osaka Mathematics Journal*, **5**, 13-52. [7.2]

Okamoto, Masashi (1963), An asymptotic expansion for the distribution of the linear discriminant function, *Annals of Mathematical Statistics*, **34**, 1286-1301. (Correction, **39** (1968), 1358-1359.) [6.6]

Okamoto, Masashi (1973), Distinctness of the eigenvalues of a quadratic form in a multivariate sample, *Annals of Statistics*, **1**, 763-765. [13.2]

Olkin, Ingram, and S. N. Roy (1954), On multivariate distribution theory, *Annals of Mathematical Statistics*, **25**, 329-339. [7.2]

Olson, C. L. (1974), Comparative robustness of six tests in multivariate analysis of variance, *Journal of the American Statistical Association*, **69**, 894-908. [8.6]

Pearl, Judea (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge. [15.1]

Pearson, E. S., and H. O. Hartley (1972), *Biometrika Tables for Statisticians*, Vol. II, Cambridge (England), Published for the Biometrika Trustees at the University Press. [Preface, 8.4]

Pearson, E. S., and S. S. Wilks (1933), Methods of statistical analysis appropriate for k samples of two variables, *Biometrika*, **25**, 353-378. [10.5]

Pearson, K. (1896), Mathematical contributions to the theory of evolution-III. Regression, heredity and panmixia, *Philosophical Transactions of the Royal Society of London, Series A*, **187**, 253-318. [2.5, 3.2]

Pearson, K. (1900), On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philosophical Magazine*, **50** (fifth series), 157-175. [3.3]

Pearson, K. (1901), On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*,**2** (sixth series), 559-572. [11.2]

Pearson, K. (1930), *Tables for Statisticians and Biometricians*, Part I (3rd ed.), Cambridge University, Cambridge. [2.3]

Pearson, K. (1931), *Tables for, Statisticians and Biometricians*, Part II. Cambridge University, Cambridge. [2.3, 6.8]

Perlman, M. D. (1980), Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations, *Annals of Statistics*, **8**, 247-263. [10.2, 10.3]

Perlman, M. D., and I. Olkin (1980), Unbiasedness of invariant tests for MANOVA and other multivariate problems, *Annals of Statistics*, **8**, 1326-1341. [8.10]

Phillips, P. C. B. (1980), The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables, *Econometrica*, **48**, 861-878. [12.7]

Phillips, P. C. B. (1982), A new approach to small sample theory, unpublished, Cowles Foundation for Research in Economics, Yale University. [12.7]

Pillai, K. C. S. (1954), On some distribution problems in multivariate analysis. Mimeo Series No. 54, Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina. [8.6]

Pillai, K. C. S. (1955), Some new test criteria in multivariate analysis, *Annals of Mathematical Statistics*, **26**, 117-121. [8.6]

Pillai, K. C. S. (1956), On the distribution of the largest or the smallest root of a matrix in multivariate analysis, *Biometrika*, **43**, 122-127. [8.6]

Pillai, K. C. S. (1960), *Statistical Tables for Tests of Multivariate Hypotheses*. Statistical Center, University of the Philippines, Manila. [8.6]

Pillai, K. C. S. (1964), On the moments of elementary symmetric functions of the roots of two matrices, *Annals of Mathematical Statics*, **35**, 1704-1712. [8.6]

Pillai, K. C. S. (1965), On the distribution of the largest characteristic root of a matrix in multivariate analysis, *Biometrika*, **52**, 405-412. [8.6]

Pillai, K. C. S. (1967), Upper percentage points of the largest root of a matrix in multivariate analysis, *Biometroka*, **54**, 189-194. [8.6]

Pillai, K. C. S., and A. K. Gupta (1969), On the exact distribution of Wilks' criterion. *Biometrika*, **56**, 109-118. [8.4]

Pillai, K. C. S., and Y. S. Hsu (1979), Exact robustness studies of the test of independence based on four multivariate criteria and their distribution problems under violations, *Annals of the Institute of Statistical Mathematics*, **31**, Part A, 85-101. [8.6]

Pillai, K. C. S., and K. Jayachandran (1967), Power comparisons of tests of two multivariate hypotheses based on four criteria, *Biometrika*, **54**, 195-210. [8.6]

Pillai, K. C. S., and K. Jayachandran (1970), On the exact distribution of Pillai's $V^{(s)}$ criterion, *Journal of the American Statistical Association*, **65**, 447-454. [8.6]

Pillai, K. C. S., and T. A. Mijares (1959), On the moments of the trace of a matrix and approximations to its distribution, *Annals of Mathematical Statistics*, **30**, 1135-1140. [8.6]

Pillai, K. C. S., and B. N. Nagarsenker (1971), On the distribution of the sphericity test criterion in classical and complex normal populations having unknown covariance matrices, *Annals of Mathematical Statistics*, **42**, 764-767. [10.7]

Pillai, K. C. S., and P. Samson, Jr. (1959), On Hotelling's generalization of T^2, *Biometrika*, **46**, 160-168. [8.6]

Pillai, K. C. S., and T. Sugiyama (1969), Non-central distributions of the largest latent roots of three matrices in multivariate analysis. *Annals of the Institute of Statistical Mathematics*, **21**, 321-327. [8.6]

Pillai, K. C. S., and D. L. Young(1971), On the exact distribution of Hotelling's generalized T_0^2,

Journal of Multivariate Analysis, **1**, 90-107. [8.6]

Plana, G. A. A. (1813), Mémoire sur divers problémes de probabilité, *Mémoires de l'Académie Impériale de Turin, pour les Années 1811-1812*, **20**, 355-408. [1.2]

Pólya, G. (1949), Remarks on computing the probability integral in orle and two dimensions, *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), 63-78. [2.P]

Rao, C. R. (1948a), The utilization of multiple measurements in problems of biological classification, *Journal of the Royal Statistical Society B*, **10**, 159-193. [6.9]

Rao, C. R. (1948b), Tests of significance in multivariate analysis, *Biometrika*, **35**, 58-79. [5.3]

Rao, C. Radhakrishna (1951), An asymptotic expansion of the distribution of Wilks's criterion, *Bulletin of the International Statistical Institute*, **33**, Part 2, 177-180. [8.5]

Rao, C. R. (1952), *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, New York. [12.5]

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), John Wiley & Sons, New York. [4.2]

Rasch, G. (1948), A functional equation for Wishart's distribution, *Annals of Mathematical Statistics*, **19**, 262-266. [7.2]

Reiersϕl, Olav (1950), On the identifiability of parameters in Thurstone's multiple factor analysis, *Psychometrika*, **15**, 121-149. [14.2]

Reinsel, G. C., and R. P. Velu (1998), *Multivariate Reduced-rank Regression*, Springer, New York. [12.7]

Richardson, D. H. (1968), The exact distribution of a structural coefficient estimator, *Journal of the American Statistical Association*, **63**, 1214-1226. [12.7]

Rothenberg, Thomas J. (1977), Edgeworth expansions for multivariate test statistics, IP-255, Center for Research in Management Science, University of California, Berkeley. [8.6]

Roy, S. N. (1939), *p*-statistics or some generalisations in analysis of variance appropriate to multivariate problems, *Sankhyā*, **4**, 381-396. [13.2]

Roy, S. N. (1945), The individual sampling distribution of the maximum, the minimum and any intermediate of the *p*-statistics on the null-hypothesis, *Sankhyā*, **7**, 133-158. [8.6]

Roy, S. N. (1953), On a heuristic method of test construction and its use in multivariate analysis, *Annals of Mathematical Statistics*, **24**, 220-238. [8.6, 10.6]

Roy, S. N. (1957), *Some Aspects of Multivariate Analysis*, John Wiley & Sons, New York. [8.6, 10.6, 10.8]

Ruben, Harold (1966), Some new results on the distribution of he sample correlation coefficient, *Journal of the Royal Statistical Society B*, **28**, 513-525. [4.2]

Rubin, Donald B., and Dorothy T. Thayer (1982), EM algorithms for ML factor analysis, *Psyhometrika* **47**, 69-76. [14.3]

Ryan, D. A. J., J. J. Hubert, E. M. Carter, J. B. Sprague, and J, Parrot (1992), A reduced-rank multivariate regression approach to joint toxicity experiments, *Biometrics*, **48**, 155-162. [12.7]

Šalaevskiĭ, O. V. (1968), The minimax character of Hotelling's T^2 test (Russian), *Doklady Akademii Nauk SSSR*, **180**, 1048-1050. [5.6]

Šalaevskiĭ, (Shalaevskii), O. V. (1971), Minimax character of ttotelling's T^2 test. I. *Investigations in Classical Problems of Probability Theory and Mathematical Statistics*, V. M. Kalinin and

O. V. Shalaevskii (Seminar in Mathematics. V. I. Steklov Institute, Leningrad, Vol. 13), Consultants Bureau, New York. [5.6]

Sargan, J. D., and W. M. Mikhail (1971), A general approximation to the distribution of instrumental variables estimates, *Econometrica*, **39**, 131-169. [12.7]

Sawa, Takamitsu (1969), The exact sampling distribution of ordinary least squares and two-stage least squares estimators, *Jorunal of the American Statistical Association*, **64**, 923-937. [12.7]

Schatzoff, M. (1966a), Exact distributions of Wilks's likelihood ratio criterion, *Biometrika*, **53**, 347-358. [8.4]

Schatzoff, M. (1966b), Sensitivity comparisons among tests of the general linear hypothesis, *Journal of the American Statistical Association*, **61**, 415-435. [8.6]

Scheffé, Henry (1942), On the ratio of the variances of two normal populations. *Annals of Mathematical Statistics*, **13**, 371-388. [10.2]

Scheffé, Henry (1943), On solutions of the Behrens-Fisher problem, based on the t-distribution, *Annals of Mathematical Statistics*, **14**, 35-44. [5.5]

Schmidli, H. (1996), *Reduced-rank Regression*, Physica, Berlin. [12.7]

Schuurmann, F. J., P. R. Krishnaiah, and A. K. Chattopadhyay (1975), Exact percentage points of the distribution of the trace of a multivariate beta matrix, *Journal of Statical Computation and Simulation*, 3, 331-343. [8.6]

Schuurmann, F. J., and V. B. Waikar (1973), Tables for the power function of Roy's two-sided test for testing hypothesis $\Sigma = I$ in the bivariate case, *Communications in Statistics*, **1**, 271-280. [10.8]

Schuurmann, F. J., V. B. Waikar, and P. R. Krishnaiah (1975), Percentage points of the joint distribution of the extreme roots of the random matrix $S_1(S_1 + S_2)^{-1}$, *Journal of Statistical Computation and Simulation*, **2**, 17-38. [10.6]

Schwartz, R. (1967), Admissible tests in multivariate analysis of variance. *Annals of Mathematical Statistics*, **38**, 698-710. [8.10]

Serfling, Robert J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York. [4.2]

Simaika, J. B. (1941), On an optimum property of two important statistical tests. *Biometrika*, **32**, 70-80. [5.6]

Siotani, Minoru (1980), Asymptotic approximations to the conditional distributions of the classification statistic Z and its studentized form Z^*, *Tamkang Journal of Mathematics*, **11**, 19-32. [6.6]

Siotani, M., and R. H. Wang (1975), Further expansion formulae for error rates and comparison of the W-and the Z-procedures in discriminant analysis, Technical Report No. 33, Department of Statistics, Kansas State University, Manhattan, Kansas. [6.6]

Siotani, M., and R. H. Wang (1977), Asymptotic Expansions for Error Rates and Comparison of the W-Procedure and the Z-Procedure in Discriminant Analysis, *Multivariate Analysis IV*, North-Holland, Amsterdam, 523-545. [6.6]

Sitgreaves, Rosedith (1952), On the distribution of two random matrices used in classification procedures, *Annals of Mathematical Statistics*, **23**, 263-270. [6.5]

Solari, M. E. (1969), The "maximum likelihood solution" of the problem of estimating a linear functional relationship, *Journal of the Royal Statistical Society B*, **31**, 372-375. [14.4]

Spearman, Charles (1904), "General-intelligence", objectively determined and measured, *American Journal of Psychology*, **15**, 201-293. [14.2]

Speed, T. P., and H. Kiiveri (1986), Gaussian Markov distributions over finite graphs, *Annals of Statistics*, **14**, 138-150. [15.5]

Srivastava, M. S., and C. G. Khatri (1979), *An Introduction to Multivariate Statistics*, North-Holland, New York. [10.9, 13.P]

Stein, C. (1956a), The admissibility of Hotelling's T^2-test, *Annals of Mathematical Statistics*, **27**, 616-623. [5.6]

Stein, C. (1956b), Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability* (Jerzy Neyman, ed.), Vol. I, 197-206, University of California, Berkeley. [3.5]

Stein, C. (1974), Estimation of the parameters of a multivariate normal distribution. I. Estimation of the means, Technical Report No. 63, NSF Contract GP 30711X-2, Department of Statistics, Stanford University. [3.5]

Stoica, P., and M. Viberg (1996), Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions, *IEEE Transaction Signal Processing*, **44**, 3069-3078. [12.7]

Student (W. S. Gosset) (1908), The probable error of a mean, *Biometrika*, **6**, 1-25, [3.2]

Styan, George P. H. (1990), *The Collected Papers of T. W. Anderson: 1943-1985*, John Wiley & Sons, Inc., New York. [Preface]

Subrahmaniam, Kocherlota, and Kathleen Subrahmaniam (1973), *Multivariate Analysis: A Selected and Abstracted Bibliography, 1957-1972*, Marcel Dekker, New York. [Preface]

Sugiura, Nariaki, and Hisao Nagao(1968), Unbiasedness of some test criteria for the equality of one or two covariance matrices, *Annals of Mathematical Statistics*, **39**, 1686-1692. [10.8]

Sugiyama, T. (1967), Distribution of the largest latent root and the smallest latent root of the generalized B statistic and F statistic in multivariate analysis. *Annals of Mathematical Statistics*, **38**, 1152-1159. [8.6]

Sugiyama, T., and K. Fukutomi (1966), On the distribution of the extreme characteristic roots of the matrices in multivariate analysis, *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers*, **13**. [8.6]

Sverdrup, Erling (1947), Derivation of the Wishart distribution of the second order sample moments by straightforward integration of a multiple, integral, *Skandinavisk Aktuarietidskrift*, **30**, 151-166. [7.2]

Tang, P. C. (1938), The power function of the analysis of variance tests with tables and illustrations of their use, *Statistical Research Memoirs*, **2**, 126-157. [5.4]

Theil, H. (assisted by J. S. Cramer, H. Moerman, and A. Russchen) (1961), *Economic Forecasts and Policy*, (2nd rev. ed.), North-Holland, Amsterdam. Contributions to Economic Analysis No. XV (first published 1958). [2.8]

Thomson, Godfrey H. (1934), Hotelling's method modified to give Spearman's "g" *Journal of Educational Psychology*, **25**, 366-374. [14.3]

Thomson, Godfrey H. (1951), *The Factorial Analysis of Human Ability* (5th ed.), University of London, London. [14.7]

Thurstone, L. L. (1947), *Multiple-Factor Analysis*, University of Chicago, Chicago. [14.2, 14.5]

Tsay, R. S., and G. C. Tiao (1985), Use of canonical analysis in time series model identification,

Biometrika, **72**, 299-315. [12.7]

Tukey, J. W. (1949), Dyadic anova, an analysis of variance for vectors, *Human Biology*, **21**, 65-110. [8.9]

Tyler, David E. (1981), Asymptotic inference for eigenvectors, *Annals of Statistics*, **9**, 725-745. [11.7]

Tyler, David E. (1982), Radial estimates and the test for sphericity, *Biometrika*, **69**, 429-436. [3.6]

Tyler, David E. (1983a), Robustness and efficiency properties of scatter matrices, *Biometrika*, **70**, 411-420. [3.6]

Tyler, David E. (1983b), The asymptotic distribution of principal component roots, under local alternatives to multiple roots, *Annals of Statistics*, **2**, 1232-1242. [11.7]

Tyler, David E. (1987), A distribution free M-estimator of multivariate scatter. *Annals of Statistics*, **15**, 234-251. [3.6]

Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986), Reduced rank models for multiple time series, *Biometrika*, **73**, 105-118. [12.7]

von Mises, R. (1945), On the classification of observation data into distinct groups, *Annals of Mathematical Statistics*, **16**, 68-73. [6.8]

von Neumann, J. (1937), Some matrix. inequalities and metrization of matric-space, *Tomsk University Review*, **1**, 286-300. Reprinted in *John von Neuman Collected Works* (A. H. Taub, ed.), **4** (1962), Pergamon, New York, 205-219. [A.4]

Wald, A. (1943), Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society*, **54**, 426-482. [4.2]

Wald, A. (1944), On a statistical problem arising in the classification of an individual into one of two groups. *Annals of Mathematical Statistics*, **15**, 145-162, [6.4, 6.5]

Wald, A. (1950), *Statistical Decision Functions*, John Wiley & Sons, New York. [6.2, 6.7, 8.10]

Wald, A., and R. Brookner (1941), On the distribution of Wilks' statistic for testing the independence of several groups of variates, *Annals of Mathematical Statistics*, **12**, 137-152. [8.4, 9.3]

Walker, Helen M. (1931), *Studies in the History of Statistical Method*, Williams and Wilkins, Baltimore. [1.1]

Welch, P. D., and R. S. Wimpress (1961), Two multivariate statistical computer programs and their application to the vowel recognition problem, *Journal of the Acoustical Society of America*, **33**, 426-434. [6.10]

Wermuth, N. (1980), Linear recursive equations, covariance selection and path analysis, *Journal of the American Statistical Association*, **75**, 963-972. [15.5]

Whittaker, E. T., and G. N. Watson (1943), *A Course of Modern Analysis*, Cambridge University, Cambridge. [8.5]

Whittaker, Joe (1990), *Graphical Models in Applied Multivariate Statistics*, John Wiley & Sons, Inc., Chichester. [15.1]

Wiisman, Robert A. (1979), Constructing all smallest simultaneous confidence sets in a given class, with applications to MANOVA, *Annals of Statistics*, **7**, 1003-1018. [8.7]

Wijsman, Robert A. (1980), Smallest simultaneous confidence sets with applications in multivariate analysis, *Multivariate Analysis*, **V**, 483-498. [8.7]

Wilkinson. James Hardy (1965). *The Algebraic Eigenvalue Problem*, Clarendon, Oxford. [11.4]

Wilkinson, J. H., and C. Reinsch (1971), Linear Algebra, Springer-Verlag, New York. [11.4]

Wilks, S. S. (1932), Certain generalizations in the analysis of variance, *Biometrika*, **24**, 471-494. [7.5, 8.3, 10.4]

Wilks, S. S. (1934), Moment-generating operators for determinants of product moments in samples from a normal system, *Annals of Mathematics*, **35**, 312-340. [8.3]

Wilks, S. S. (1935), On the independence of k sets of normally distributed statistical variables, *Econometrica*, **3**, 309-326. [8.4, 9.3, 9.P]

Wishart, John (1928), The generalised product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A**, 32-52. [7.2]

Wishart, John (1948), Proofs of the distribution law of the second order moment statistics, *Biometrika*, **35**, 55-57. [7.2]

Wishart, John, and M. S. Bartlett (1933), The generalised product moment distribution in a normal system, *Proceedings of the Cambridge Philosophical Society*, **29**, 260-270. [7.2]

Wold, H. D. A. (1954), Causality and econometrics, *Econometrica* **22**, 162-177. [15.1]

Wold, H. D. A. (1960), A generalization of casual chain models, *Econometrica*, **28**, 443-463. [15.1]

Woltz, W. G., W. A. Reid, and W. E. Colwell (1948), Sugar and nicotine in cured bright tobacco as related to mineral element composition, *Proceedings of the Soil Sciences Society of America*, **13**, 385-387. [8.P]

Wright, Sewall (1921), Correlation and causation, *Journal of Agricultural Research*, **20**, 557-585. [15.1]

Wright, Sewall (1934), The method of path coefficients, *Annals of Mathematical Statistics*, **5**, 161-215. [15.1]

Yamauti, Ziro (1977), *Concise Statistical Tables*, Japanese Standards Association. [Preface]

Yates, F., and W. G. Cochran (1938), The analysis of groups of experiments, *Journal of Agricultural Science*, **28**, 556. [8.9]

Yule, G. U. (1897a), On the significance of Bravais' formulae for regression & c., in the case of skew correlation, *Proceedings of the Royal Society of London*, **60**, 477-489. [2.5]

Yule, G. U. (1897b), On the theory of correlation, *Journal of the Royal Statistical Society*, **60**, 812-854. [2.5]

Zehna, P. W. (1966), Invariance of maximum likelihood estimators, *Annals of Mathematical Statistics*, **37**, 744. [3.2]